Synthetic Incident Generation in the Reenactment of IT Support Organization Behavior

Claudio Bartolini
HP Labs
Hewlett Packard
Palo Alto, CA, USA
claudio.bartolini@hp.com

Cesare Stefanelli, Mauro Tortonesi
Department of Engineering
University of Ferrara
Ferrara, Italy
{cesare.stefanelli,mauro.tortonesi}@unife.it

***Abstract*—In IT service management, IT support organizations are the entities in charge of restoring normal service operation after a disruption. Building accurate models of IT support organizations is useful for several purposes, such as optimal workforce allocation and what-if scenario analysis. However, the complexity of real-life IT support organizations makes it extremely hard to model their organizational structure and their behavior with stochastic processes. A particularly interesting process to model in the reenactment of IT support organization is the incident arrival one. The paper presents three synthetic incident generator models, based on advanced statistical methods, that are capable of reenacting, with different levels of accuracy, the incident arrival process of real-life IT support organizations. The methods were developed from the experience that the authors developed in the experimental analysis of transaction logs from a real-life IT support organization, provided by the Outsourcing Services Division of HP.**

IT support organizations.

I. Introduction

IT support organizations have a primary role in IT service management. Among other things, they are in charge of incident management, which ITIL defines as the process for restoring normal service operation after a disruption [1] [2]. Real-life IT support organizations usually implement very complex organizational, structural, and behavioral processes according to the strategic objectives defined at the business management level.

Researchers have built accurate theoretical models of IT support organizations for purposes such as optimal workforce allocation [3] or shift scheduling to guarantee low waiting times in incident servicing [4]. In this context, a particularly interesting application scenario is the modeling of IT support organization for *what-if scenario analysis*.

What-if scenario analysis is a technique based on the construction of an accurate model of the system under evaluation and on its successive exploitation to re-enact the system behavior with modified parameters. When applied to IT support organizations, what-if scenario analysis is an instrument to evaluate alternative incident management strategies before putting them in practice, therefore significantly improving the efficiency of a very expensive and time-consuming process.

To apply what-if scenario analysis in IT support organizations, we developed SYMIAN, a decision support tool for the performance analysis and optimization of the IT incident management function. SYMIAN re-enacts the behavior of IT support organizations using an open queuing network model and discrete event simulation. SYMIAN has been used to enable the optimization of IT support organizations according to both performance-driven [5] and business impact-driven criteria [6].

During our experiments with SYMIAN, we have been able to reproduce the behavior of real-life IT support organizations with good enough fidelity. More specifically, we used a stationary Poisson process to model incident arrival and GI/M/s-FCFS first-come-first-served multi-server queues to model support groups [5]. However, while SYMIAN's fidelity was acceptable to a first order approximation, the tool did not accurately reproduce some minor aspects of IT support organizations, such as queue lengths, and sojourn time (the time an incident resides at a specific support group, including waiting and service time) distribution. This motivated us to keep refining our model, thereby increasing its accuracy.

This paper focuses on one of the main issues that we have to deal with in order to take our simulation-based approach to the next level, and specifically on the modeling of the incident arrival process. In fact, our experience demonstrates that the accurate reenactment of the incident arrivals is essential in order to develop more sophisticated solutions for the performance improvement of IT support organizations

In our experience, we have observed that many of the most commonly used distributions, such as Poisson, Gamma, and power-law, do not allow a realistic synthetic generation of incident inter-arrival times [7]. So, the present paper focuses on how to derive a more accurate approximation of the incident arrival process.

In order to answer this issue, it is necessary to gather additional insights on the behavior of real-life IT support organizations. To this purpose, we performed an experimental analysis of transaction logs from a real-life IT support organization, provided to us by the Outsourcing Services Division of Hewlett-Packard. HP Outsourcing manages, among other IT services, the Help Desk function on behalf of various enterprise customers. The data used for this experiment comes from the subset of the organization serving a single enterprise customer from the financial services industry, whose name will be disguised as BailUsOut in the remainder of the paper.

The statistical analysis of transactional logs from the BailUsOut IT support organization allowed us to develop 3 synthetic incident generator models to reenact the arrival process, which we named GPD, NHPP, and WPAR. The models are based on different approaches that build on top of statistical analysis techniques, and respectively *stationary parametric probability distribution functions*, *non-homogeneous Poisson processes*, and *ad hoc resampling of observed data*.

We experimentally validated the synthetic incident generator models, by using them as inputs for the BailUsOut IT support organization model we built in SYMIAN. We compared the simulation outcomes, and found that the WPAR model captures the behavior of the incident arrival process with the highest degree of accuracy.

While the BailUsOut IT support organization might not be representative of every IT support organization, the lessons learned from its analysis are very important. In fact, they allow us to build a more accurate model of the organization with SYMIAN, as well as to gain useful experience in the modeling process.

II. IT Support Organizations

IT support organizations are typically composed of a network of support groups, each employing a set of operators, with potentially different work shift schedule. Support groups are divided into support levels (usually three to five), with lower level groups dealing with generic issues and higher level groups handling technical and time-consuming tasks. Support groups are further specialized by category of incidents that they deal with (network, server, etc…) and usually organized by geography, to ensure prompt incident response.

The Help Desk provides the lowest support level and represents the interface for customers reporting an IT service disruption. In response to a customer request, the Help Desk "opens" an incident, sometimes also called trouble-ticket or simply ticket. The incident is then "assigned" to a specific support group, whose technicians either fully repair the incident or "reassign" it to a different support group (usually escalating to a higher support level).

As a result, an incident goes through different states and is handled by different support groups throughout its lifetime. At each of these steps, the incident record is updated with the pertinent information, such as current state and related service restoration activity. If, for some reason, customers request the organization to stop working on the incident, the incident is placed in a "suspended" state to avoid incurring into Service Level Objective penalties. Once the disruption is repaired, the ticket is placed in "closed" state until the end-user confirms that the service has been fully restored. In this case, the incident is "resolved" and its lifecycle ends.

The characteristics of IT support organizations suggest to model them as open queuing networks [8], as it is done for telephone call centers [9] [3]. However, notice that IT support organizations have some peculiar characteristics that distinguish them from telephone call centers. In fact, in IT support organizations, there is no need to consider call blockings, abandonments, or redials. Instead, there is the need to consider routing of incidents through the system as well as complex incident prioritization policies since IT support organizations might serve many customers with different profiles, each one with a specific SLA.

III. Modeling the Incident Arrival Process

A particularly important aspect in modeling IT support organization is represented by the incident arrival process. In fact, the accurate modeling of the incident arrival process is of critical importance for realistic simulations.

Re-enacting incident traces from real IT support organization is an often-used idea to increase the accuracy of simulation outcomes. However, this limits the applicability of the simulation to predefined input, thus preventing its use to verify how the modeled organization would behave under circumstances such as heavy incident load or a specific inter-arrival pattern. Synthetic incident generation according to configurable stochastic patterns is essential to overcome this limitation.

In order to develop synthetic incident generator models for IT support organizations, the most promising approaches are represented by 3 different statistical analysis techniques: *stationary parametric probability distribution functions*, *non-homogeneous Poisson processes*, and *ad hoc resampling of observed data*.

Unfortunately, in our experience we have found that the nature of incident arrival processes in real-life IT support organizations is essentially non-stationary and changes over time. This suggests that stochastic models based on stationary parametric probability distribution functions, often used in similar applications, might not be appropriate to model the incident arrival process in IT support organizations.

Instead, the non-stationary behavior of the incident arrival process could be better captured using stochastic processes with time-varying intensities, such as non-homogeneous Poisson processes [11] or Poisson-Gamma processes [12], which are often proposed for similar applications.

However, our experience demonstrates that the incident arrival process in real-life IT support organizations is not only time-varying, but also very complex. As a result, techniques based on multiple non-stationary processes and non-parametric statistical techniques [13], such as ad hoc resampling of observed data, while significantly more complex, represent a more promising approach to model the incident arrival process.

The next Sections discuss the 3 synthetic incident generation models we developed, one for each of the above mentioned categories, and present an experimental evaluation of the synthetic incident generators in the context of the BailUsOut IT support organization model that we built in SYMIAN.

IV. Synthetic Incident Generation for the BailUsOut IT Support Organization

We analyzed inter-arrival times at the BailUsOut IT support organization. Fig. 1 shows a histogram that represents the inter-arrival time distribution for incidents. We observe that most of the incident inter-arrival times are in the order of one minute, but some of them are very high. As a result, the inter-arrival time distribution is L-shaped, with a long tail.

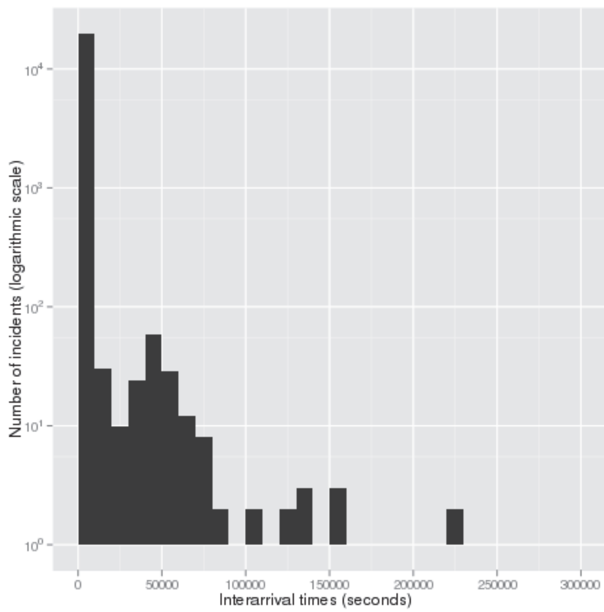

Figure 1. Histogram of incident interarrival times.

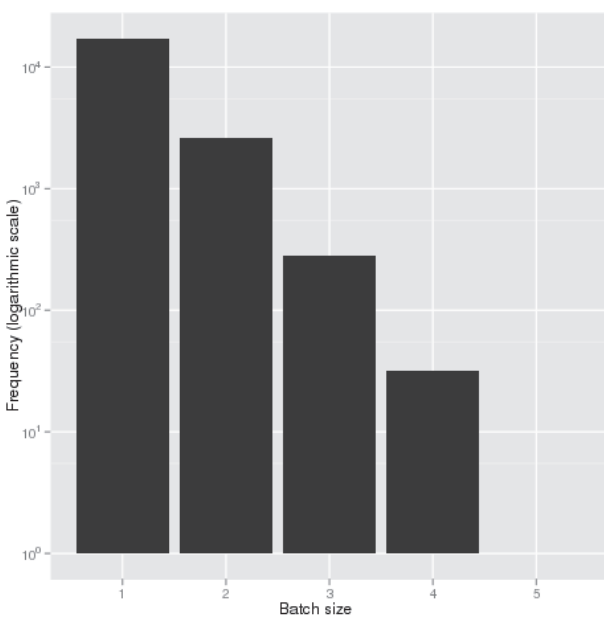

Figure 2. Histogram of incident batch arrival sizes.

We also observed that a significant portion of incidents (26.39%) arrive in batches. Fig. 2 shows a histogram that represents the distribution of batch sizes. We found that the discrete logarithmic probability distribution is a rough, although not statistically valid, approximation to the batch size distribution.

A. *Model 1: General Pareto Distribution (GPD)*

While we were not able to match the distribution of the incident interarrival times in the BailUsOut IT support organization against any common distribution function, we have found that power law distributions represent the closest approximation.

Therefore, we developed a model of incident arrivals based on the adoption of a general Pareto probability distribution function to capture the incident interarrival process. The general Pareto distribution is often used to model stochastic processes of the long-tail type, for instance in applications where extreme events need to be considered. Because of its flexibility, the general Pareto distribution lends itself to a wide range of applications. In fact, it represent the best choice, among the most well-known parametric probability distribution functions, to capture the behavior of the incident management process in IT support organizations.

To evaluate the capability of the GPD model to reenact the incident arrival process, we have applied it to create a synthetic incident generator for the BailUsOut IT support organization. More specifically, in order to infer the model parameters we have applied a maximum likelihood procedure, based on a vector generalized linear model [14], to the incidents arrivals observed in the BailUsOut dataset. To this end, we have used the functions provided by the VGAM package [15] for the R statistical framework [20]. Assuming a location value of 0, the inference process returned a scale value of 136.89 (minutes) and a shape value of 0.52607 (minutes).

Fig. 3 compares the outcome of our GPD synthetic incident generator model with the data from transactional logs. As one can see, the level of accuracy of this model is not very high. The model coarsely reproduces the historical incident arrival process, generating incidents with smoother interarrival time values. More specifically, the model generates a larger number of incidents with very low or high interarrival times but does not capture the behaviour of the incident arrival process with regard to incident with average interarrival times. In addition, this model does not enable to generate batch incident arrivals.

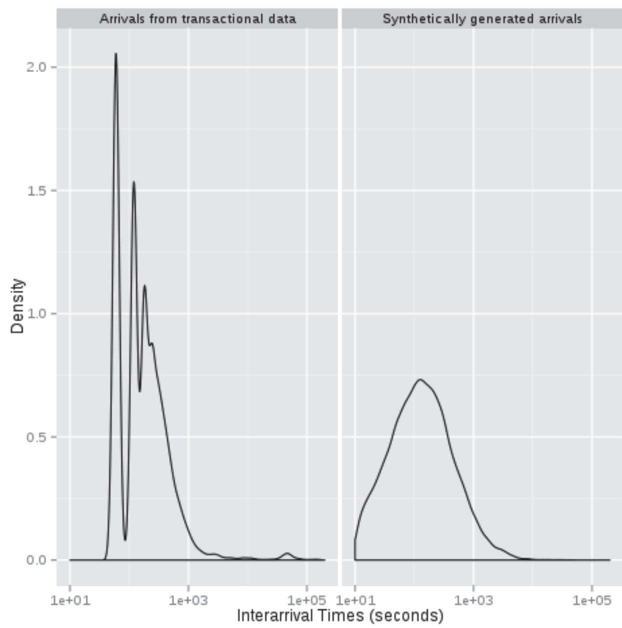

Figure 3. Comparison of kernel density estimation for the incident traces extracted from transactional data and produced by the GPD synthetic incident generator model.

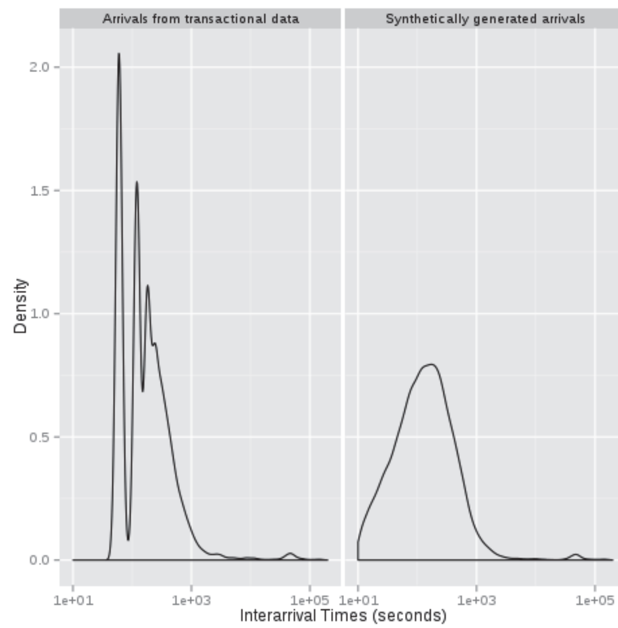

Figure 4. Comparison of kernel density estimation for the incident traces extracted from transactional data and produced by the NHPP synthetic incident generator model.

B. Model 2: Non-Homogenous Poisson Process (NHPP)

Following the approach most commonly found in literature for the modelling of non-stationary stochastic processes, we developed a second synthetic incident generation model based on non-homogeneous Poisson processes.

Non-homogeneous Poisson processes are a well known and widely used method to model non-stationary stochastic processes. More specifically, non-homogeneous Poisson processes are an extension of simple stationary Poisson processes where the rate parameter λ is not a constant, but a function $\lambda(t)$ that changes over time.

We have created a synthetic incident generator for the BailUsOut IT support organization using the NHPP model, in order to evaluate its capability to reenact the incident arrival process in real-life IT support organizations. More specifically, to simplify the model parameter inference process, we assumed that the arrival rate function $\lambda(t)$ remains constants within 1-hour time intervals. We then analyzed transactional logs to obtain the value that $\lambda(t)$ assumes in each of the above mentioned intervals.

Fig. 4 compares the outcome of our NHPP synthetic incident generator model with the data from transactional logs. Like the GPD model, the NHPP model has a rather low level of accuracy and tends to generate a larger portion of incidents with very low or high interarrival times. The NHPP model also does not enable to generate batch incident arrivals.

C. Model 3: Weekly Pattern Aware Resampling (WPAR)

Finally, we developed a third model of incident arrivals based on ad hoc resampling of observed data. This model is based on the observation that the nature of the arrival process is complex and changes fundamentally over time, and that in order to accurately capture it multiple non-stationary processes and non-parametric statistical techniques are required.

More specifically, we created the WPAR model according to the transactional log analysis that we ran on the BailUsOut dataset, and we later applied the model to create a synthetic incident generator for the corresponding IT support organization.

Since we observed that the arrival process is stationary on weekly averages, we speculated that the cause of this L-shaped distribution in incident inter-arrivals originated in the differences between weekdays and non-working (weekend) days and we proceeded to investigate that hypothesis. Fig. 5 shows a classification of incidents, according to week day of arrival and time to next incident arrival. As one can see, short inter-arrival times (less than one hour) are distributed only during weekdays, medium inter-arrival times (1 to 24 hours) are evenly distributed across all week, and long inter-arrival times (24 hours or more) almost exclusively appear on non-working days.

This analysis demonstrates that the incident arrival process exhibits a dynamical behavior with a 7-day periodicity and explains why stationary probability distribution functions cannot accurately capture incident arrival dynamics.

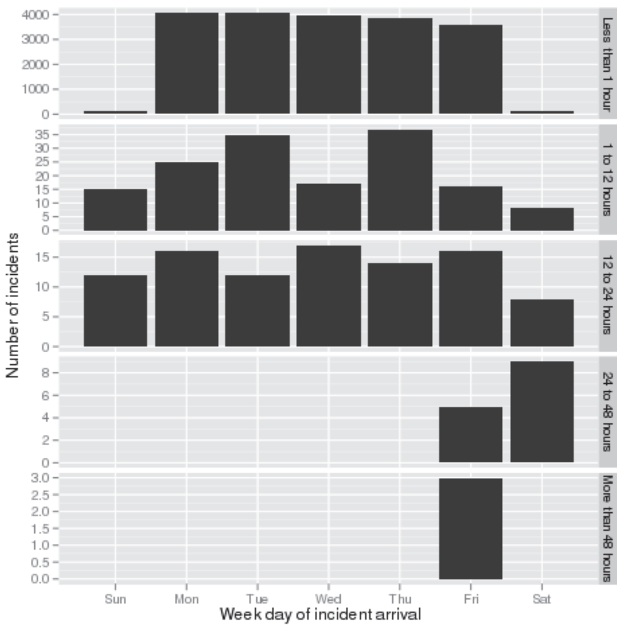

Figure 5. Histogram of incident arrivals classified by week day of arrival (on the X axis) and time to next arrival (on the Y axis at the right).

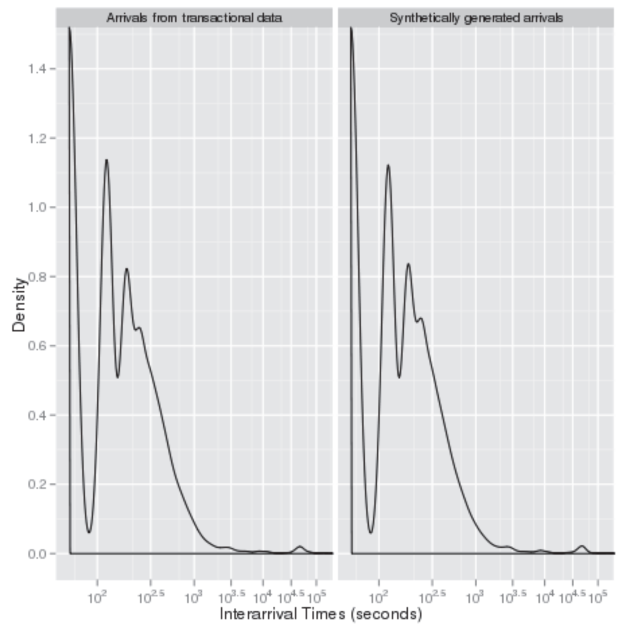

Figure 6. Comparison of kernel density estimation for the incident traces extracted from transactional data and produced by the WPAR synthetic incident generator model.

Since we have found almost no correlation (the Pearson sample correlation coefficient is -0.01101) between arrival times and batch sizes, we can model incident arrivals and batch size independently as separated processes.

Using the information collected in the previous analysis, we were able to develop a highly accurate synthetic incident generator. Since no parametric distribution could yield a good approximation of the arrival process observed in the BailUsOut IT support organization, we leveraged on non-parametric statistical methods. For maximum accuracy, we also adopted a time-varying incident generation approach.

More specifically, we separated the data in two different sets: one containing incidents arrived on weekdays and the other containing incidents arrived on weekends. For the first set, we calculated incident inter-arrival times by considering as a valid timeframe only weekdays, i.e., the inter-arrival time between incident A arrived on a Friday at 11:59 PM and incident B arrived on the next Monday at 0:14 AM is 15 minutes. In a similar way, we calculated incident inter-arrival times for the second set of incidents by considering only weekends. For each set, we then obtained the empirical cumulative distribution functions for both inter-arrival time and batch size.

We then developed a synthetic incident generation algorithm that, according to the current day of the week in simulated time, picks the time to next incident arrival and the size of next incident batch by sampling from the corresponding empirical cumulative distribution functions. The outcome of our synthetic incident generator approximates the data from transactional logs very accurately (Fig. 6).

We also performed a Wilcoxon sum rank test [13] on the incident trace produced by the synthetic incident generator and the one derived from transactional data, to verify if they come from the same probability distribution function. The high p-value we obtained (0.9648) does not allow us to reject the null hypothesis and suggests that the samples are equally distributed.

These result demonstrate that the WPAR model is capable of re-enacting the BailUsOut IT support organization incident arrival process with a significantly higher accuracy than the GPD and the NHPP models.

V. EXPERIMENTAL EVALUATION

We used the synthetic incident generation models described in the previous Section as inputs for the BailUsOut IT support organization. More specifically, we used the BailUsOut IT support organization model based on GI/M/s-FCFS queues that we developed in [5].

We ran 4 different simulations, respectively using as input the historical incident trace and each of the GPD, NHPP, and WPAR synthetic incident generator models. Each simulation covering 6 months of simulated time, plus a warm-up time of 15 days, introduced in order to prime the simulation environment. The warm-up phase is necessary in order not to take unrealistic measurements on a cold start. In this phase, events such as incident arrivals, closures, and escalations are discarded and not taken into account for the evaluation of the organization performance metrics.

Fig. 7 shows the number of incidents arrived to the BailUsOut IT support organization model. As one can see, the number of incidents generated by the synthetic incident generator models is rather close to the number of incidents in the historical transactional log trace. However, the high number of incidents generated by the GPD has a non-negligible impact on the system-wide behavior of the IT support organization, as we will see shortly.

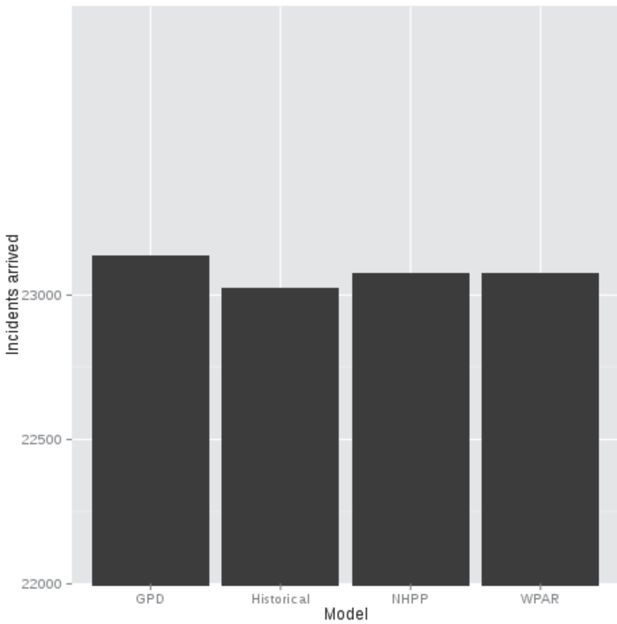

Figure 7. Comparison of arrived incidents using input from historical traces and from each of the models introduced in Section IV.

In fact, Fig.8 shows the number of incidents closed, i.e., fully resolved, by the BailUsOut IT support organization model. In this case, the GPD synthetic incident generator model performs much worse than the NHPP and WPAR models. This is due to the larger number of incidents that the GPD model generates. The incident generation rate of the GPD model outpaces the IT support organization capability of dealing with incidents, thereby slowing down the entire system.

Finally, Fig. 9 shows a comparison of the Mean Time To Resolution (MTTR) metric (expressed in minutes), that captures the average response time that the system exhibits when dealing with full incident restoration. Clearly, the GPD model is severely outperformed by the NHPP and WPAR models with regards to the capability to realistically capture the historical trace behavior. Again, this is due to the larger number of incidents that the GPD model generates. The WPAR generator is more accurate than the NHPP one, and represents the synthetic incident model with the best overall performance.

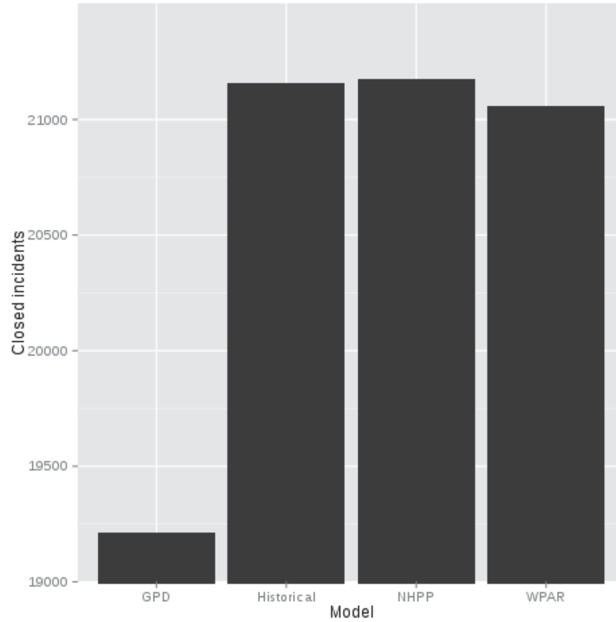

Figure 8. Comparison of closed incidents using input from historical traces and from each of the models introduced in Section IV.

The results presented in Fig. 8 and Fig. 9 clearly demonstrate the need to consider an accurate incident generation model when reenacting IT support organizations.

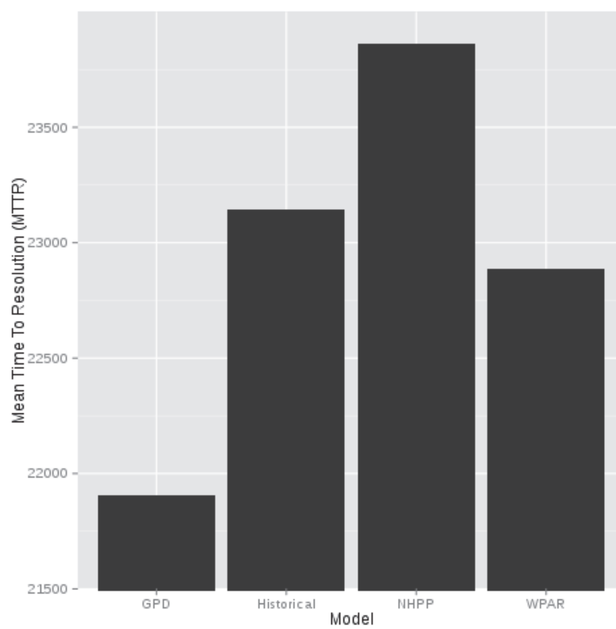

Figure 9. Comparison of MTTR (in minutes) using input from historical traces and from each of the models introduced in Section IV.

Notice that the IT support organization models based on GI/M/s-FCFS queues, such as the BailUsOut model used in the context of this experimental evaluation, are not perfectly capable of capturing high frequency dynamics, such as those exhibited at the single support group level, in the IT support organization behavior. We expect that the adoption of the GPD, NHPP, and WPAR in the context of more sophisticated IT support organization models, for instance based on multiple-priority queues [10], might lead to even larger differences.

VI. Related Work

There is an extensive research literature on queuing models for support organizations. For a representative example, see [9] and [3].

Most of the previous research has focused on telephone call centers. Brown et al. perform an analysis of real-life telephone call center transaction logs, observing non-stationary behaviors in the arrival process similar to what we find in our BailUsOut dataset [12]. However, they focus on the statistical analysis of parameters, such as customer arrival intensity, that are of interest for the optimization of workforce allocation in order to improve quality of service. Instead, we are interested in the development of a model for the reenactment of an entire IT support organization to adopt in what-if scenario analysis.

Only recently have researchers started studying IT support organizations. See for example the works by Wasserkrug et al. [4], focusing on shift scheduling for IT support personnel and reporting on the inadequacy of M/M/N queues to model support groups, and by Zeltyn et al. [11], developing analytic models for multi-server multi-priority queues with applications in IT support organizations. Our work adopts a different approach, as we are interested in the development a simulative model of real-life IT support organizations starting from the analysis of transactional logs.

A particularly interesting field of research involves data mining over ticket information. Medem et al. have applied document clustering techniques [17], also in correlation with information provided by automated incident reporting tools [16], to ticket description data. Their analysis of semantic incident attributes represents a step towards a deeper understanding of how IT support organizations deal with tickets, and is complementary to the transactional log analysis presented in this paper.

The EasyTicket system [18] is an interesting machine learning application which aims at optimizing the ticket routing in IT support organizations considering the incident routing history. In a more recent publication, the same authors propose the Optimized Network Model, which also considers the tickets' content [19]. However, while automated routing suggestion tools are certainly valuable, it is still unclear whether they can outperform routing decisions made by expert personnel. In fact, it might be as well be the opposite, since automated routing suggestion systems need to be trained using a dataset that might contain obsolete or inaccurate information. In addition, these systems usually do not consider (variations in) the state of support group, e.g., to prefer routing to support group with lower workload in case multiple routing decisions are available. Finally, our experience suggests that in real-life IT support organization social networking effects may emerge that can possibly improve the performance of human routing decisions, e.g., operator A from support group X becomes friend with operator B from support group Y and uses that contact to quickly route tickets to their resolution.

To the best of our knowledge, there are no other papers that examine transactional data to draw conclusions on modeling incident arrival process in IT support organizations.

VII. Conclusions and Future Work

We demonstrated that the incident arrival process can be accurately reproduced using our WPAR model. However, it is possible that in some cases the WPAR synthetic incident generator might produce input traces that are too specific. Other applications, instead, might require even more accurate synthetic incident generators, which also consider time of day in incident arrival modeling. This calls for the careful evaluation of the synthetic incident generation model to adopt in what-if scenario analysis according to the specific application, and also raises some interesting questions about the tradeoffs of model complexity – that unfortunately cannot be addressed here.

Future work will attempt to extend the WPAR model to enable the reshaping of the generated incident trace, to increase the fit to the input data trace and/or to simulate the effect of workload increase (or decrease). In addition, we would like to evaluate the use of different synthetic incident generation models in combination.

Finally, it has to be noted that the models described above rest on an assumption of incident independence. In reality, incidents are symptoms of underlying problems, and therefore, incidents related to the same problem are not independent. However, the excellent accuracy of the synthetic incident generators presented in this paper makes us lean against explicitly considering dependencies of incidents onto problems, which would significantly increase the model complexity.

References

[1] Office of Government Commerce: *Service Strategy Book*. The Stationery Office, Norwich (2007)

[2] Office of Government Commerce: *Service Operation Book*. The Stationery Office, Norwich (2007)

[3] Koole, K.: "Call Center Mathematics". VU University of Amsterdam, Department of Mathematics Website. http://www.math.vu.nl/~koole/ccmath/. Accessed 30 August 2010.

[4] Wasserkrug, S., Taub, S., Zeltyn, S., Gilat, D., Lipets, V., Feldman, Z., Mandelbaum, A.: "Creating operational shift schedules for third-level IT support: challenges, models and case study". *Intern. J. of Serv. Oper. and Inform*., 3(3-4), 242-257 (2008)

[5] Bartolini, C., Stefanelli, C., Tortonesi, M.: "SYMIAN: Analysis and Performance Improvement of the IT Incident Management Process". *IEEE Trans. on Netw. and Serv. Manag.*, 7(3) (2010)

[6] Bartolini, C., Stefanelli, C., Tortonesi, M.: "Business-impact analysis and simulation of critical incidents in IT service management". In *Proc. 11th IFIP/IEEE Intern. Symp. on Integr. Netw. Manag.* (2009)

[7] Bartolini, C., Stefanelli, C., Tortonesi, M.: "Modeling IT Support Organizations from Transactional Logs". In *Proc. 12th IEEE/IFIP Netw. Oper. and Manag. Symp.* (2010)

[8] Bolch, G., Greiner, S., de Meer, H., Trivedi, K.: *Queuing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications* (2nd Ed.). Wiley, Hoboken (2006)

[9] Koole, G., Mandelbaum, A.: "Queueing Models of Call Centers: An Introduction". *Ann. of Oper. Res.*, 113(1-4), 41-59 (2002)

[10] Bartolini, C., Stefanelli, C., Tortonesi, M.: "Modeling IT Support Organizations Using Multiple-Priority Queues". In *Proc. 13th IEEE/IFIP Netw. Oper. and Manag. Symp.* (2012)

[11] Zeltyn, S., Feldman, Z., Wasserkrug, S.: "Waiting and sojourn times in a multi-server queue with mixed priorities". *Queueing Syst.*, 61(4), 305-328 (2009)

[12] Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., Zhao, L.: "Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective". *J. of the Am. Stat. Assoc.*, 100(469), 36-50 (2005)

[13] Kvam, P., Vidakovic, B.: *Nonparametric Statistics with Applications to Science and Engineering*. Wiley, Hoboken (2007)

[14] Yee, T. W., Wild, C. J., "Vector Generalized Additive Models", *J. of the Royal Statist. Soc., Series B, Methodological*, 58(3), 481-493 (1996)

[15] Yee, T. W.: "The VGAM Package for Categorical Data Analysis", *J. Statist. Soft.*, 32(10), 1-34 (2010)

[16] Medem, A., Teixeira, R., Feamster, N., Meulle, M.: "Joint Analysis of Network Incidents and Intradomain Routing Changes". *In Proc. 6th IFIP/IEEE Intern. Conf. on Netw. And Serv. Manag.* (2010)

[17] Medem, A., Akodjenou, M.-I., Teixeira, R.: "TroubleMiner: Mining network trouble tickets". In *Proc. 11th IFIP/IEEE Intern. Symp. on Integr. Netw. Manag.* (2009)

[18] Shao, Q., Chen, Y., Tao, S., Yan, X., Anerousis, N.: "Efficient Ticket Routing by Resolution Sequence Mining". In *Proc. 14th ACM SIGKDD Intern. Conf. on Knowl. Discov. and Data Min.* (2008)

[19] Miao, G., Moser, L., Yan, X., Tao, S., Chen, Y., Anerousis, N., "Generative Models for Ticket Resolution in Expert Networks". In *Proc. 16th ACM SIGKDD Intern. Conf. on Knowl. Discov. and Data Min.* (2010)

[20] R Development Core Team: "R: A Language and Environment for Statistical Computing", R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org/ (2011)