Towards the Knowledge-Driven Benchmarking of Autonomic Communications

David Lewis, Declan O'Sullivan, John Keeney

Knowledge and Data Engineering Group, Trinity College Dublin

{Dave.Lewis\Declan.OSullivan\John.Keeney}@cs.tcd.ie

Abstract

Currently a wide range of different adaptive and intelligent system solutions are being proposed for use in self-managing or autonomic networks. However, there are few means by which such proposals can be compared. This paper proposes that a benchmark be developed for autonomic systems so that progress in this field can be more systematically evaluated. Our approach assumes that autonomic systems make use of and thus expose a knowledge based representation of the service they offer, the context they react to and the governance to which they are subject. This position paper focuses on some of the issues that arise when formulating a benchmark for autonomic communications and is intended to form the basis for further discussion in the area.

1. Introduction

In the last few years there has been intense activity in the study of so called 'autonomic systems'. This originally emerged to support the operation of complex computing systems [1], but has rapidly been adopted by subject by researchers in communications [2] and pervasive computing [3]. The term autonomic is in many places used synonymously with the idea of self-management, in particular where self management refers to self-configuring, self-healing, self-optimising and self-protecting behaviour [1]. However, though many examples exist of systems claiming to exhibit one or more of these behaviours a commonly accepted means for comparing these systems is not available. This stems from a wider ambiguity about exactly what does and does not constitute an autonomic system.

In terms of agreeing a common understanding of what constitutes an autonomic system, it is widely acknowledged that an autonomic system is not a fully automatic system insofar as it must be subject to human governance at some point, typically specified in terms of high-level governance policies. In general these policy specifications are more closely aligned to human operational goals than to the system management terms used in the deployed policy-based management systems today.

In this paper we specifically examine how a benchmark could be developed for autonomic communication systems, i.e. communication networks that exhibit self-managing behaviour. We first examine initial discussion on benchmarking from the Autonomic Computing community. We then examine the particular issues related to the benchmarking of Autonomic Communication, and examine existing benchmarking approaches in the communications industry. Finally we propose an abstract model that can be used as the basis for autonomic benchmarking and discuss some of the benchmarks that might be amenable to early agreement.

2. Autonomic Computing Benchmarks

A primary goal of an autonomic system is to dramatically reduce the cost of operating a system. As computing and communication systems become more complex, the cost of operating these systems has been observed to grow disproportionately. These costs grow from the need for administrative staff to understand the complexity of the system in order to administer it. This includes the cost of mal-administration resulting from performing administrative tasks with an incomplete understanding of their system-wide ramifications. Autonomic systems aim to deliver this cost reduction by relieving human administrators of some of the cognitive load associated with administering complex systems. The autonomic function consists of a closed control loop, from which the human administrator is removed. The human administrator is instead able to govern the operation of this control loop by specifying high-level policies, which are used to determine strategies for how the system should self-manage.

IBM's widely recognised reference architecture for autonomic computing [4] describes this control loop as linking the monitoring of system state and context, analyzing this monitored data to support the planning of adjusted operational behaviour which is executed through actuators to control the systems behaviour. A feature of this control loop is that knowledge rich metadata is used to communicate between the parts of the control loop, enabling flexible interaction and the ready integration of artificial intelligence features into the analysis part loop, e.g. AI planning or Bayesian analysis. In this work we assume that knowledge-based representations are used to expose the semantics of the sensed information an autonomic system monitors, the service it provides to its users and the model via which it is governed.

Though this control loop provides an abstraction for discussing the architecture of various approaches to implementing autonomic systems, it does not provide us with a clear means to assessing and thus comparing the level of autonomicity exhibited by different systems. Ultimately, the level of autonomicity of a system has to be assessed in a holistic manner. For instance if the introduction of a self-configuring feature into a system results in more complex human administration of security, the reduction in overall operational cost will be adversely effected. This should thus impact on the measure of marginal autonomicity provided by the introduction of the feature.
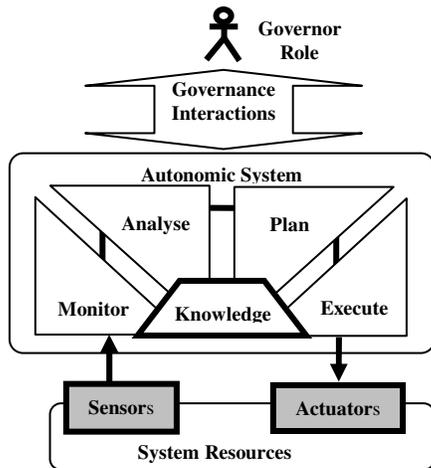

Figure 1: The autonomic process

Considering the issue of a holistic assessment of autonomic features more broadly, we must assess the impact of introducing an autonomic feature on the whole lifecycle cost of the system. For instance the feature must ensure that the user's experience of the system is not adversely affected by its introduction. Further, we must assess the impact of new autonomic feature on the maintenance and upgrade costs of the system. In other words the introduction of autonomic features must not prove a hindrance to further innovation of the system. Therefore, any approach to benchmarking autonomic systems must be framed by the question of whether the increased complexity and cost inherent in introducing an autonomic feature is outweighed by the reduction in both system operation cost and in total cost of ownership.

Currently there is no accepted framework for calculating the operational cost via which an autonomic system may be assessed. IBM define a scale for describing how autonomic the management of a system is in terms of how much human intervention is required for its management, namely: Basic, Managed, Predictive, Adaptive and Autonomic [5], but at present no structured criteria are available to measure this human input in a general purpose manner.

In one of the few articles published on autonomic system benchmarks, Brown et al discuss the implication of extending existing performance benchmarks to handle autonomic systems [6]. This requires benchmarks that inject changes into the system in addition to the quiescent load typically applied in benchmarks. It is the system's ability to cope with these changes with as little intervention as possible from the human administrator that forms the basis for their proposed autonomic benchmark. This presented several unresolved challenges in terms of making the injection of changes repeatable and, more significantly, making them representative of real world change, e.g. the release of a network worm, or major damage inflicted on the physical infrastructure.

Brown et. al. also suggest a set of dimensions that should characterize the response of an autonomic system to injected change:

- Level of response to injected change
- Quality of response to injected change
- Impact of response on human users
- Cost of extra resources needed to support the autonomic response

Here the operational cost is affected by the level of response and the quality of response to the injected changes, these being assessed primarily by the difference in workload and cognitive load imposed on the system's administrator. This needs to be combined with the cost of extra resources that need to be employed in the response to the change to derive a measure of the total cost of the autonomic feature. In

the ideal situation the impact on the user should be zero, however, where a negative impact occurs a Service Level Agreement (SLA) may be in place to help quantify the level to which service degradation violated the level of quality expected by the user. Without an SLA, benchmarking with service degradation becomes difficult, but not impossible. In a live operational setting, the lack of an SLA may be addressed by qualitative customer satisfaction surveys, which attempt to gauge the loss of customer goodwill or trust in the service due to a perceived drop in value in the system due inadequate responses to changes in operational environment such as failures or load peaks. Here perceived changes in value may be subjective in nature, but when an autonomic communications service is considered to be a value chain of components and services, cost and value can be offset against each other within the chain to calculate value-add within the network. The aim of autonomic communications management is to increase the perceived value in the network to a greater extent than the cost of adding autonomic management.

Another attempt at outlining a benchmark for autonomic computing is presented in [7]. This proposes the following metrics, on which we comment:

Quality of Service (QoS): as discussed above usually relies on an SLA, and will tend to be highly application specific. Ideally, comparison of autonomic features can be performed against a common adherence to an SLA.

Cost: in terms of the cost of procuring the autonomic feature and the coupled lifetime costs of this addition, plus the operational cost changes related to the administration of the autonomic function concerns.

Granularity/Flexibility: on the basis that fine-grain systems will offer more scope for adaptive behaviour in response to system perturbation.

Failure avoidance/robustness: must incorporate the degree to which failures have been predicted, or whether robustness to some level of unanticipated event is evident. By definition, however, only spot checks can be conducted in response to unanticipated changes, so responses to these will be at best a probabilistic measure.

Degree of Autonomy: relates to robustness and is a measure to which the system copes with unanticipated changes in the environment. Like robustness similar issues to use as a benchmarking metric apply, though picking randomly from a large set of possible changes might provide some suitable mechanism.

Adaptivity: can be interpreted as the breadth of different types of changes to which the system can response. In our approach this can be characterised by

the semantic spread of the context information to which the system is know to react.

Time to react and adapt: are the time to detect that a change required adaptive behaviour and the time taken for the system to effect a change needed to handle the change without impacting the QoS experience by users, or at least not beyond the bounds of the SLA. With a well controlled means of injecting changes into a benchmarked system, this should be relatively easy to measure.

Stability: is the time taken by the system to learn its environment and stabilise its operations. This is important since the mean of adaptation incurs resource usage (management messages, CPU cycles), so the longer it runs the longer these resources may be unavailable for service delivery. This is related to adoption time, but with benchmarks that inject multiple changes to reflect system subject to continuous environmental flux.

Sensitivity: captures the level of environmental change to which the system reacts. This is not so much a measure of autonomic quality, but a tuning parameter – too high and the system may become unstable, too low and it looses adaptivity.

Although the proposed benchmark from McCann describes the characteristics of an autonomic system, many of these characteristics cannot be completely defined or measured in a meaningful general-purpose manner. However, many of the concepts can be captured to some degree using representative metrics. From this we can see that while some interesting features of an autonomic system can be measured via benchmarks that inject perturbations into the systems environment, some features are really design facets used to predict reaction to unanticipated event.

Currently there seems little attempt to establish a benchmarking system for the governance interface to autonomic system, despite that fact the cognitive load incurred by policy-based management interface will be a big determinant in the human cost of governing a system. Broadly, a difference in cognition and workload is acknowledged between action policies, goal policies and utility policies [8].

3. Issues in Communications Benchmarking

Why measure the autonomic aspect of autonomic communications systems?

- To measure autonomicity service levels at runtime, for use in governance policies and service

agreements, but also for use by the system itself as a runtime metric to drive management in a feedback manner to control the stable emergence of desired network-wide behaviour

- To scientifically evaluate autonomicity in general, and so compare and stimulate research efforts, thus providing a principled grounding for progressing the scientific discipline of autonomics.
- To demonstrate and communicate progress in achieving autonomicity specifically when applied to communications management.
- To compare autonomic systems for procurement purposes.

Autonomic communications studies how individually self managing network elements self-adjust in a manner constrained by high-level governance, and how this affects the operation of the element, other elements forming groups, and the end-to-end operation of the network. If the goal of Autonomic Communication research is to understand how desired element's behaviours are learned, influenced or changed, and how, in turn, these effect other elements, groups and network [2], then the degree to which elements, groups, and network are successful in achieving this goal must be measurable.

Traditional benchmarks focus on achieving quality of service, but in general these QoS metrics do not even correctly measure the customers' satisfaction with the overall quality of the service presented but rather some characteristic metric or key performance indicator of the systems used for the presentation of the service. In general, though aspects of the network operation, e.g. ease of use of OSS applications, are measured, the autonomicity of the network in terms of cost of ownership, excellence in customer experience or value perception and facilitation of new services and technologies, are not measured. This is further exacerbated by multiple definitions of what concepts or metrics contribute to a measurement of the autonomicity of a system, e.g. there are no agreed interpretations for concepts such as dependability, sensitivity, stability, complexity, etc. The establishment of common terms and definitions to describe and measure key aspects of the operation of autonomic elements, groups, and networks would assist in the progression of the research area.

In coming up with a benchmarking framework for autonomic communication, it is important to recognise that the communication industry already has fairly mature schemes for assessing operational system quality, for instance the TL9000 scheme which is based on ISO9000 approaches to quality measurements and has become a industry agreed approach to communication service quality [9]. To date standardised general purpose evaluation metrics for small-scale self-managing networks, e.g., networking in the home, have not emerged.

4. Autonomic Benchmarking Concepts

Here we lay out some specific concepts in our knowledge driven approach to a benchmarking autonomic systems. The first assertion is that autonomic communication systems should be modelled in a way that clearly distinguished between the service offered to the user, the governance that drives the self-management of the system and how the system monitors environmental context and its model of itself. Although not strictly defined, here context should be considered to encompass the situation and behaviour of the entire system, including the set of users and their relative preferences [10]. Secondly, we assume that ontological models are used in expressing these distinct views of a system. We have described such models in [11] and [12] and have identified that suitable ontological languages exist for defining service interfaces, e.g. WSMO, OWL-S. Here, we examine the potential benefits of ontological models for the model of governance that a system offers and the context view it uses. We do not, however, prescribe here what form these models should take, this being a subject of active research in the policy-based management community [13], the context aware system community [14] and the network management community [15].

To estimate the costs of the ontology-based governance space offered by an autonomic communication system, will require measurements related to the human aspects of ontology use and policy specification maintenance. The cost related to ontology use depends significantly on the ability of the user to understand the ontology, which is typically influenced by two factors: the complexity and the clarity of the ontology. Bontas et al. propose an *Ontology Understandability* cost as a combination of an ontology complexity cost driver and a ontology clarity cost driver [16]. In addition, they make some suggestions as to how these costs can be automatically determined. Although the users and governors are unlikely to be exposed to raw ontologies but rather some abstracted information model, the cost of understanding and using this model must be considered. The costs related to policy specification maintenance will decrease significantly in situations where the user works frequently with particular policy primitives. Therefore an estimation of the user's likely

familiarity with particular policy specification sets will influence the cost involved with maintenance. Similarly some measure of the complexity involved in the modification of the policy is needed. Finally, the capabilities and the experience of the users also need to be measured as these will have impact both on the costs of ontology understanding and policy maintenance.

In order to expose the amount of heterogeneity involved in the context space that needs to be handled by an autonomic system, some semantic measure would be useful. Three types of semantic measures have been proposed in the literature. The first, *semantic similarity*, evaluates the resemblance between two concepts from a subset of significant semantic links (e.g. is-a). *Semantic relatedness* evaluates the closeness between concepts from the whole set of their semantic links. Finally *semantic distance* evaluates the disaffection between two concepts, in essence an inverse notion of semantic relatedness. Of these measures, calculating the semantic distances between the ontological concepts used in the context information would be a useful indicator of the extent of heterogeneity involved. A range of semantic distance measures have been proposed, such as [17][18][19][20].

5. Summary and Further Work

Here we provide the basis for a common approach to how Autonomic Communication systems can be evaluated and benchmarked. The specification and definition of key metrics such as system lifetime cost, service lifetime cost, complexity, quality, dependability, resilience, adaptability, sensitivity etc. will be a key part of this work. This will be particularly difficult in the scope of Autonomic Communications since the communication systems are made up of heterogeneous distributed devices, agents and services, each of which will be autonomically self managing and self adjusting to some extent. To demonstrate this approach, it has already been applied to the design of a benchmark for an autonomic knowledge delivery network [21]. Preliminary use and evaluation of this benchmark is ongoing.

The degree to which autonomic communications can be benchmarked is dependent on the degree to which all of the key activities of the system can be measured and benchmarked. This will require that context-awareness, policy refinement, monitoring and interpreting, learning and planning, etc, are each evaluated and benchmarked in a concise but complete manner. As advances in each of these areas progress in

a clearly evaluated manner, a better understanding of Autonomic Communications will emerge.

However, such research will require a high level of sustained cooperation and coordination across a critical mass of stakeholders in the international telecommunications and research communities.

In order to perform an objective measurement of autonomicity in a dynamic communications system, both the context envelope and the quality of the service requested and presented will need to be constrained within a target QoS profile for the duration of the measurement. This will necessitate the use of a managed testbed to allow independent variables to be controlled, while the autonomicity aspects of the system are rigorously measured. Such a testbed could also be used to determine the levels of autonomicity or situational awareness required from individual network elements and group to facilitate the autonomic management of the entire communications system. Specific objectives in a roadmap to an autonomic communication benchmarking scheme are:

- Define concepts and terms and gain industry agreement
- Determine which metrics of an autonomic communication system best capture the autonomic qualities and quality of service characteristics of the system
- Develop a demonstrator testbed for running autonomic communication benchmarks.
- Define a framework and methodology for the consistent evaluation of these metrics in a manner that autonomic communication systems and services can be compared and benchmarked

Acknowledgements

References

[1] Kephart, J., Chess, D., "The Vision of Autonomic Computing", IEEE Computer, Jan 2003, pp 41-50
[2] M. Smirnov, "Autonomic Communication: Research Agenda for a New Communications Paradigm", Fraunhofer FOKUS, November 2004. (http://www.fokus.gmd.de/web-dokumente/Flyer_engl/Autonomic-Communicatin.pdf).
[3] A. Ranganathan, R. Campell, "Autonomic Pervasive Computing based on Planning" in proc of the 1st Int'l

Conference on Autonomic Computing, May 17-18 2004, New York, USA, pp 80-87

[4] "An architectural blueprint for autonomic computing". IBM whitepaper 2004

[5] A. G. Ganek T. A. Corbi, "The dawning of the autonomic computing era", IBM Systems Journal, 42(1), 2003, pp. 5-19.

[6] A. Brown, J. Hellerstein, M. Hogstrom, T. Lau, S. Lighthouse, P.Shum, M. Yost, "Benchmarking Autonomic Capabilities: Promises and Pitfalls", in proc of the 1st Int'l Conference on Autonomic Computing, May 17-18 2004, New York, USA, pp 266-268

[7] J. A. McCann and M. C. Huebscher. "Evaluation issues in autonomic computing". In Proceedings of Grid and Cooperative Computing Workshops (GCC), LNCS 3252, 597-608. Wuhan, China. October 21-24, 2004.

[8] Kephart, J., Walsh, W., "An Artificial Intelligence Perspective on Autonomic Computing Policies", in Proceedings of 5th IEEE International Workshop on Policies and Distributed Systems and Networks, IEEE, 2004, pp 3-12

[9] TL 9000 Quality System Metrics, book two, release 2.5, Quest Forum

[10] M. Bazire, P. Brézillon, "Understanding Context Before Using It", in Proceedings of CONTEXT 2005. 2005. Paris, France.

[11] D. Lewis, O. Conlan, D. O'Sullivan, V. Wade, "Managing Adaptive Pervasive Computing using Knowledge-based Service Integration and Rule-based Behavior", in Proceedings IFIP/IEEE Network Operations and Management Systems, Seoul Korea, 19-23 April 2004, pp 901-902

[12] J. Keeney, K. Carey, D. Lewis, D. O'Sullivan, V. Wade, "Ontology-based Semantics for Composable Autonomic Elements", Workshop of AI in Autonomic Communications at the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland 30th July 2005

[13] G. Tonti, J. M. Bradshaw, R. Jeffers, R. Montanari, N. Suri1, A. Uszok, "Semantic Web Languages for Policy Representation and Reasoning: A Comparison of KAoS, Rei, and Ponder", Proceedings of 2nd International Semantic Web Conference (ISWC2003), October 20-23, 2003, Sanibel Island, Florida, USA

[14] H. Chen, F. Perich, T. Finin, A. Joshi, "SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications", International Conference on Mobile and Ubiquitous Systems: Networking and Services, August 2004.

[15] J. E. López de Vergara, V. A. Villagrá, J. Berrocal, "Semantic Management: advantages of using an ontology-based management information meta-model", Proceedings of the HP Openview University Association Ninth Plenary Workshop (HP-OVUA'2002), distributed videoconference, 11-13 June 2002

[16] E. P. Bontas, M. Mocho, "Towards a Cost Estimation Model for Ontology Engineering", Proceedings of the 3rd Berliner XML Tage, 2005.

[17] R. Rada, H. Mili, E. Bicknell, M. Blettner, "Development and application of a metric on semantic nets", IEEE Transactions on Systems, Man, and Cybernetics 19, pp 17-30, 1989.

[18] J. Jiang, D. Conrath, "Semantic Similarity based on corpus statistics and lexical taxonomy", Proceedings of International Conference on Research in Computational Linguistics, 1997.

[19] M. Sussna, "Word sense disambiguation for free-text indexing using a massive semantic network", Proceedings of the 2nd Int'l Conference on Information and Knowledge Management, pp 67-74, 1993.

[20] Presentation, NIST Invitational Workshop on Semantic Distance, Gaithersburg, MD, November 2003.

[21] J. Keeney, D. Lewis, D. O'Sullivan, "Benchmarking Knowledge-based Context Delivery Systems", to appear in Proceedings of the International Conference on Autonomic and Autonomous Systems (ICAS 06), Silicon Valley, USA, July 19-21, 2006