Demo: Explaining Automated Environments: Interrogating Scripts, Logs, and Provenance Using Voice-Assistants

Chris Norval
University of Cambridge, UK
chris.norval@cl.cam.ac.uk

Jatinder Singh
University of Cambridge, UK
jatinder.singh@cl.cam.ac.uk

ABSTRACT

The grand vision of pervasive computing and the Internet of Things (IoT) involves providing people with a range of seamless functionality, be it through automation, information delivery, etc. However, much of the IoT is opaque; it is often difficult for users to uncover and understand how and why particular functionality occurs, the sources of information, the entities involved, and so forth. We argue that automation scripts, as well as logs and provenance records could be leveraged to assist in illuminating the workings of connected and automated environments.

This paper explores the use of voice assistants (an accessible, intuitive and increasingly common interface) as a means for allowing users to interrogate what is happening in the IoT systems that surround them. In presenting an exploratory Alexa 'Skill', we discuss several considerations for the implementation of such a system. This work represents a starting point for considering how such assistants could help people better understand—and indeed, evaluate, challenge, and accept—technology that is increasingly pervading our world.

CCS CONCEPTS

• **Human-centered computing → Interaction techniques**; **Ubiquitous and mobile computing**.

KEYWORDS

explainability, transparency, provenance, automation, voice

ACM Reference Format:
Chris Norval and Jatinder Singh. 2019. Demo: Explaining Automated Environments: Interrogating Scripts, Logs, and Provenance Using Voice-Assistants. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3341162.3343802

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6869-8/19/09.
https://doi.org/10.1145/3341162.3343802

1 INTRODUCTION

Our physical environments are becoming increasingly interconnected as the grand vision of the Internet of Things (IoT) becomes a reality. Such environments entail so-called 'smart' functionality (be it through automation, information retrieval, etc.). We already see various functionality being deployed; one might have their heating automatically switch on when they come home on a cold day, for example. Services such as IFTTT,[1] Stringify,[2] Yonomi,[3] and SmartThings,[4] (to name a few) allow the creation and use of simple automation scripts (often termed *'recipes'*) to drive IoT environments. As IoT ecosystems become ever more pervasive, it will become increasingly important that users can understand the workings of their surrounding environments – including how and why certain actions occurred, and from where information is sourced. This may be particularly relevant when unexpected or surprising events occur: *"why did my door lock?!"*

Providing visibility over the sequence of events driving automated environments could help users to understand what is happening, thereby assisting issues of user empowerment, trust and adoption [1, 4, 16]. This is by giving users some insight into what led to a particular action (e.g. *"the heating was turned on as the temperature was low"*), and the information driving its occurrence (*"the temperature was received from BBC Weather"*; *"it was recorded four hours ago"*; etc.).

In more complex arrangements, technical measures that record how such systems operate (i.e. through logging their past actions and behaviours, and the scripts driving them) will be particularly important. The work of the provenance community—which has long advocated the recording and analysis of the lineage of data and its surrounding context—may be especially relevant for establishing how such records

[1]https://ifttt.com
[2]https://www.stringify.com
[3]https://www.yonomi.co
[4]https://www.smartthings.com

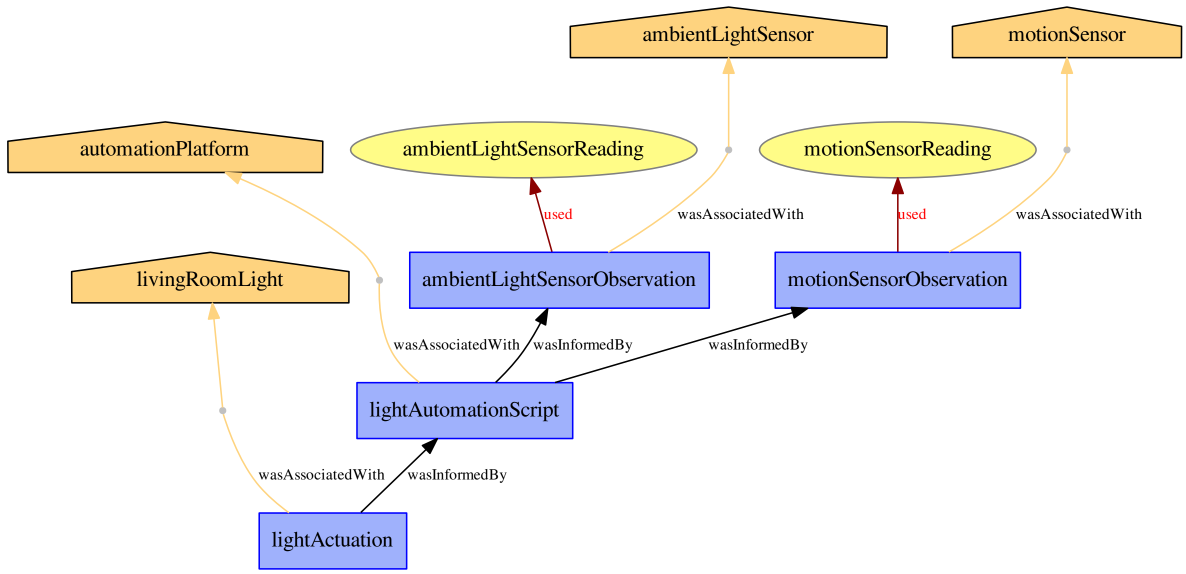

Figure 1: A PROV representation of our IoT arrangement. Here, an automation script uses the input of two sensors (a light sensor and a motion sensor) to determine whether a light should be switched on. This could represent automated lighting at a person's front door, for example.

should be captured and described. Provenance concerns the documentation of the sequence of steps (along with their inputs and outputs) of a particular system or process (e.g. see Fig. 1), with common applications including the documentation of reproducible scientific research and of complex data processing [10]. Prior research has considered audit logging [8] and provenance [17] for the IoT. Yet, even where such records are collected, the usability and utility of this information remains a challenge [2, 8, 15].

This paper explores the potential for voice assistants as a means for users to probe and interrogate the inner workings of connected and automated environments. We describe a prototype whereby the voice assistant allows users to ask questions about the functionality of systems, where answering entails leveraging and analysing the details of the underlying automation scripts, logs and provenance records. In doing so, we indicate the need for—and a step towards—more intuitive means for helping users to better understand the workings of the technical environments that they occupy.

2 EXPLANATIONS THROUGH VOICE ASSISTANTS

Our approach involves functionality within (or accessed via) the voice assistant which allows the user to ask questions about the automated environment they are in. The voice assistant then leverages recipe, log and provenance information to provide a meaningful response about system behaviour. This allows a user to query both what a system *did* do, and what a system *should* do.

We focus on voice assistants for several reasons. Firstly, the use of natural language has long been considered an intuitive way for people to interact with technology [11, 12]. Second, voice assistants are already commonplace (to date, over 100 million Amazon Alexa devices are reported to have been sold [5]). Third, they are increasingly embedded within IoT ecosystems, acting as an interface between the user and their various IoT devices [6]. Further, they are often used by their owners as a means for requesting information (e.g. *"What is the weather today?"*) [6]. As such, voice assistants represent a natural point for interrogating the happenings within an IoT environment.

Implementation

To explore the feasibility and potential of this approach, we have developed a proof-of-concept prototype. This works as follows: (i) the user asks the voice assistant a question (i.e. *"why did my door unlock?"*); (ii) the system constructs a query based on what the user is looking for; (iii) the relevant records are retrieved; (iv) the results are translated into a meaningful response; and then (v) this response is presented to the user. The user can then choose to ask follow-up questions in order to further interrogate the system, should they desire.

Information about how the connected environment *should* and/or *did* behave can be derived from recipes, log and provenance information. We represent this information in *W3C PROV*: a *"specification to express provenance records, which contain descriptions of the entities and activities involved in* *producing and delivering or otherwise influencing a given object"* [3], stored in a Neo4j graph database[5] (as is common for storing PROV representations, e.g. [9, 14]). Fig. 1 presents an example PROV arrangement of an automated environment, where two sensors control the actuation of a light. Metadata is associated with nodes to facilitate explanation, including how a node should be 'spoken about' in a user-centric way (i.e. the language and words to use).

Our implementation uses Amazon Alexa as the voice assistant platform through the creation of a *Skill* (think: 'app') which provides the means for speech-based interrogation. This Skill acts as an interface between the user and the PROV data, constructing a query based on the user's request, and querying the Neo4j database. Note that the concept is agnostic to the technology stack (the platforms were selected due to their prominence), and similar implementations could be created to work with other voice assistants, databases, or data representations.

3 TOWARDS MEANINGFUL INTERROGATION

Our prototype implementation provides a starting point as to how automation scripts and records of systems events can be leveraged to provide an answer or explanation to the user. The primary consideration is *what* information should be returned to the user. We now explore some considerations regarding meaningful user interactions.

Relevant Information: The Scope of the Response

One can envisage situations where the user seeks a 'broad' overview of the system; perhaps they require a general understanding of how certain happenings relate, or because they seek some context before they start interrogating further. For such circumstances, our Skill can describe all (or parts) of the graph of (relevant) events driving system functionality, as appropriate. Given that such orchestrations typically comprise a directed acyclic graph [13], a starting point is to identify a relevant 'root' node of the graph, and then to iteratively traverse the relevant subsequent nodes and relay their relationships with each other to the user. Describing the example of Fig. 1, for instance, entails nine such relationships. Of course, the actual utility provided by this approach, which entails explaining the whole graph, will depend on the situation. Reciting all nodes via speech may not be particularly helpful, especially for large and complex representations, though may be suitable in some simple scenarios.

In other situations, a user may be interested in what drove a particular happening, rather than a more general explanation. For example, knowing that the temperature triggered an action, the user might then inquire: *"Where did this temperature come from?"* Towards this, a 'localised' approach

[5]https://neo4j.com

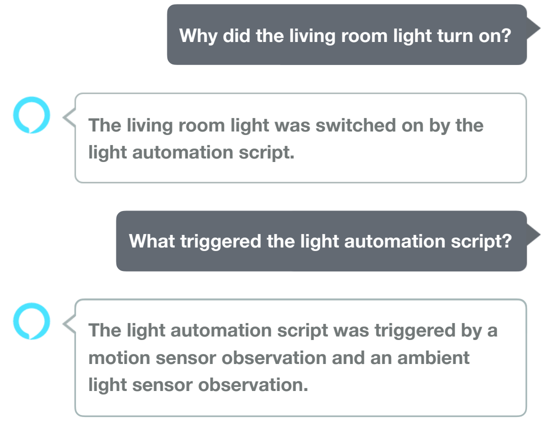

Figure 2: A localised inquiry. This demonstrates how the user could explore and traverse the records by having an initial point of interest and asking follow-up questions.

allows a user to specify their interest; in effect, to request information about a particular object or action, which results in the system querying the Neo4j database for the relevant node. The node of interest can then be returned (spoken aloud), along with its relevant relationships with directly subsequent nodes. The user may then go on to ask about one of those nodes; effectively acting to traverse the graph of relevant events step-by-step, should they so require (Fig. 2).

Interpreting User Questions

Language and NLP research will be relevant in interpreting user requests, allowing more natural and granular questioning, (e.g. *"Is this temperature value an actual reading or a forecast?"*), which works to inform and construct more relevant queries. Ontologies and standards will also play an important role. There appears interesting research opportunities regarding how language can best be used to help explain systems, particularly where the type of language could be tailored to user characteristics and expectations, (e.g. age, experience, technical expertise). Also relevant are means for enabling better conversational abilities – a recognised limitation of the current generation of voice assistants [7].

Adapting the Response

There may be situations where it is more suitable to provide a response to the user through another modality, rather than verbally; for instance, if the response is complex, or contains sensitive information. A visual response, for example, could involve delivering the answer to a 'second screen', be it one on the assistant itself (if it has one), or the user's phone, laptop, television, etc. This could be in the form of a textual description, visualisation (e.g. as per Fig. 1, or a comic [15]), or some hybrid approach that combines various modalities to best explain the environment to a given user.

There is room for devising new interaction representations and modalities, and investigating their suitability in different situations through user studies. Naturally, the most appropriate method for answering user queries will depend on the context, and so approaches that tailor responses accordingly is likely a key area for research. Nevertheless, *speech will often be an appropriate query interface*, even where the results/ output may better suit a different form of presentation.

Further, there may well be different audiences for the information returned; the background of the user (and their expectations) will often be relevant. For instance, what is deemed appropriate for a non-technical user may be different to that of a regulator auditing the connected environment (both in terms of the nature of the information, and how it is returned) [16]. Such aspects also require consideration.

The Need for Systems Data

All that which has been discussed hinges on the availability of data regarding how the system operates – which is often not collected or generally accessible. As such, we argue that more needs to be done regarding the (i) collection, (ii) processing, and (iii) availability of information regarding the functionality of connected and automated environments. Such information will provide the basis for enabling the development of interaction methods for making this information more meaningful to users, and other relevant parties (e.g. auditors and regulators). More broadly, our work seeks to inspire the exploration of what might be possible in how such data can be used to better explain complex systems.

4 CONCLUDING REMARKS

In all, there are increasing societal demands for giving users greater transparency regarding the technical systems affecting them, especially as computing becomes more pervasive. Beyond the speech context, by demonstrating a concrete and tangible method for systems interrogation, we seek to help motivate and encourage increased attention by the community towards tools and methods that work to better empower users, thereby assisting users in navigating, probing, interrogating and understanding complex systems arrangements.

ACKNOWLEDGMENTS

This work was supported by the UK EPSRC, and Microsoft through the Microsoft Cloud Computing Research Centre.

REFERENCES

[1] Muhammad Naveed Aman, Kee Chaing Chua, and Biplab Sikdar. 2017. Secure Data Provenance for the Internet of Things. In *Proceedings of the 3rd ACM International Workshop on IoT Privacy, Trust, and Security (IoTPTS '17)*. ACM, New York, NY, USA, 11–14.

[2] Khaled Bachour, Richard Wetzel, Martin Flintham, Trung Dong Huynh, Tom Rodden, and Luc Moreau. 2015. Provenance for the People: An HCI Perspective on the W3C PROV Standard Through an Online Game. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2437–2446.

[3] Khalid Belhajjame, Helena Deus, Daniel Garijo, Graham Klyne, Paolo Missier, Stian Soiland-Reyes, and Stephen Zednik. 2013. PROV Model Primer. https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/. Accessed: 2019-01-30.

[4] Stanislav Beran, Edoardo Pignotti, and Peter Edwards. 2015. Interrogating Capabilities of IoT Devices. In *Provenance and Annotation of Data and Processes*, Bertram Ludäscher and Beth Plale (Eds.). Springer International Publishing, Cham, 197–202.

[5] Dieter Bohn. 2019. Amazon Says 100 Million Alexa Devices Have Been Sold: — What's Next? https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp. Accessed: 2019-07-12.

[6] BT. 2018. BT Modern Families Report. https://img01.products.bt.co.uk/content/dam/bt/storefront/pdfs/BT-Modern-Families-Report_20nov.pdf. Accessed: 2019-07-12.

[7] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. ACM, New York, NY, USA, Article 43, 12 pages.

[8] Andy Crabtree, Tom Lodge, James Colley, Chris Greenhalgh, Kevin Glover, Hamed Haddadi, Yousef Amar, Richard Mortier, Qi Li, John Moore, Liang Wang, Poonam Yadav, Jianxin Zhao, Anthony Brown, Lachlan Urquhart, and Derek McAuley. 2018. Building accountability into the Internet of Things: the IoT Databox model. *Journal of Reliable Intelligent Environments* 4, 1 (01 Apr 2018), 39–55.

[9] Hugo Firth and Paolo Missier. 2015. ProvGen: Generating Synthetic PROV Graphs with Predictable Structure. In *Provenance and Annotation of Data and Processes*, Bertram Ludäscher and Beth Plale (Eds.). Springer International Publishing, Cham, 16–27.

[10] Melanie Herschel, Ralf Diestelkämper, and Houssem Ben Lahmar. 2017. A survey on provenance: What for? What form? What from? *The VLDB Journal* 26, 6 (01 Dec 2017), 881–906.

[11] Matthew B. Hoy. 2018. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly* 37, 1 (2018), 81–88.

[12] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5286–5297.

[13] Luc Moreau, Paul Groth, Simon Miles, Javier Vazquez-Salceda, John Ibbotson, Sheng Jiang, Steve Munroe, Omer Rana, Andreas Schreiber, Victor Tan, and Laszlo Varga. 2008. The Provenance of Electronic Data. *Commun. ACM* 51, 4 (April 2008), 52–58.

[14] E. Nwafor, A. Campbell, D. Hill, and G. Bloom. 2017. Towards a provenance collection framework for Internet of Things devices. In *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation.* 1–6.

[15] Andreas Schreiber and Regina Struminski. 2017. Visualizing Provenance using Comics. In *9th USENIX Workshop on the Theory and Practice of Provenance (TaPP 2017)*. USENIX Association.

[16] J. Singh, J. Cobbe, and C. Norval. 2019. Decision Provenance: Harnessing Data Flow for Accountable Systems. *IEEE Access* 7 (2019), 6562–6574.

[17] Qi Wang, Wajih Ul Hassan, Adam Bates, and Carl Gunter. 2018. Fear and Logging in the Internet of Things. In *Network and Distributed Systems Symposium.*