Aretha: A Respectful Voice Assistant for the Smart Home

William Seymour, Max Van Kleek, Reuben Binns, Nigel Shadbolt

Department of Computer Science, University of Oxford, Oxford, UK, OX1 3QD
{william.seymour, max.van.kleek, reuben.binns, nigel.shadbolt} @ cs.ox.ac.uk

Keywords: respect, voice assistants, smart home, IoT, internet of things

Abstract

Despite being novel and convenient, voice assistants have brought with them a myriad of privacy and security related concerns. Previous research has shown how the ubiquitous nature of data collection combined with the lack of controls available to users can lead to apathy and dejected acceptance of the status quo. In this paper we present the design of Aretha, a speculative voice assistant that radically shifts the power balance in the smart home. Aretha is able to have conversations about privacy and security with users, helping them to change and adapt their preferences over time. These preferences can then be enforced using network-level controls, effectively retrofitting good behaviour to existing devices.

1 Introduction

The development of voice assistants such as the Amazon Echo and Google Home has rapidly transformed the promise of the 'smart home' from science fiction to science fact. But as these devices have grown in popularity, society is starting to become aware of the problems presented by voice assistants, and scholars have become increasingly concerned about the privacy implications of devices designed (in many cases) to advertise and surveil their users.

At the same time, human factors research into online privacy and security reveals that while users do express concern over these risks, they are typically unlikely to act on those concerns. This occurs for a myriad of complex reasons, including users lacking the required knowledge to identify problems and the fact that individual preferences are much more complex than previously thought.

The ability for voice assistants in their role as hubs to observe the behaviour of other IoT devices opens the door for solutions that use this position not to gather data for marketing or data collection purposes, but to act as a watchdog for other devices. This paper presents the design of Aretha, a prototype voice assistant created to improve user awareness and proficiency around smart home security and privacy.

Making use of offline analytics and modern voice recognition technology, Aretha is able to bring something to the smart home that has been missing for a long time: *conversations* about security and privacy. These conversations help scaffold users' understanding of what is happening on their home networks, helping them to develop a situational awareness of the data flows in their home. Aretha then encourages users to follow up and act on this understanding through a series of powerful and easy to use enforcement mechanisms.

In the following sections, we demonstrate how shifting the design calculus of the voice assistant to one that truly works for its *users*, rather than one that prioritises the interests and conveniences of manufacturer and/or third parties, could impact technological and interpersonal interactions in the home. In addition to presenting a technical overview of the Aretha software and hardware stacks, we describe how Aretha:

1. Helps scaffold *understanding* about security and privacy issues by providing accessible analysis of smart device network traffic
2. Utilises voice conversations and the Socratic method to develop this understanding, helping users form privacy and security *preferences*
3. by providing options linked to (1) and (2), significantly lowers the barrier to entry for acting on security and privacy concerns through the provision of powerful and easy to use *enforcement mechanisms*

2 Background

2.1 Privacy and Security in the Smart Home

Many attempts to design products and interventions to communicate security and privacy risks are borne of the long-running narrative in privacy research that users express concern over privacy and security, yet consistently do not take privacy and security preserving actions. Acquisti et al. decompose the problem, showing how legal and regulatory structures (such as privacy policies) often place users in an impossible situation when navigating privacy choices [1], suggesting that users make choices armed with incomplete information; that even

with perfect information, they are unlikely to be able to process all of it; and that even with perfect information and the capacity to process it, human psychological deviations often cause us to deviate from rationality.

This is a sentiment echoed by many other contributions. Users are broadly held to know little about how their information is used [2], and feel that they have way to change it even if they do know how it is used [3]. This has led to suggestions that future solutions should support users by making suggestions and taking actions on their behalf, rather than simply providing more information [4].

Brunton and Nissenbaum similarly highlight the deliberately vague terminology in privacy policies, as well as the 'fantasy of opting out', a situation complicated by the introduction of the European General Data Protection Regulation (GDPR). The legal nature of these policies necessitates "rendering who is doing what puzzling and unclear in terms of justifications", suggesting a forced disconnect between the imagined relationship the user believes they have with manufacturers of their devices and the real, legal one defined by nebulous terms and conditions [5].

While examples of open source voice assistants do exist, such as Mycroft[1] and the now defunct Jasper assistant[2], as well as software more geared towards automation, such as Home Assistant[3], none of these existing projects promote user understanding in the way that Aretha does—Mycroft, for example, provides an open platform, but it does not help users to understand how it or other devices send data to remote parties.

2.2 Modelling Privacy

The question of how to model informational privacy is a source of intense disagreement. One of the most widely accepted models of privacy is Nissenbaum's theory of privacy as contextual integrity, based on norms surrounding the appropriateness of information in context, and flows of information between different parties [6]. In this way, the appropriateness of a data flow is determined by the combination of sender, receiver (*actors*), the *context* in which the information is shared, the *norms* that exist around information sharing in that context, *attributes* of the shared information, and the *principles* of the transmission.

For example, in a healthcare context it is normal for a patient's medical information to be shared with their doctor, with the understanding that this information is kept in confidence. A colleague knowing the same information would constitute a violation of contextual integrity; it is not the sharing of medical records or sharing with colleagues *per se* that is wrong, but the combination of actor, context, and the other factors described above.

This helps to explain why smart devices (and voice assistants in particular) are so aptly positioned to violate their users' privacy. Their position in the home allows them to observe users in many different contexts and collect various different types of information, and they often have poorly defined transmission principles (i.e. users do not know what data is collected by these devices or how it is processed); even benign devices can easily become responsible for inappropriately filtering and propagating information across boundaries, as they lack the required contextual understanding of the environment in which they operate. Aretha is therefore positioned as a network level 'manager' of these information flows, (literally) conversing with users to elicit preferences and interceding on their behalf when inappropriate flows are discovered.

2.3 The Social Nature of Voice Interfaces

Voice assistants occupy a unique position amongst smart home devices, owing to the personal nature of the data they collect coupled with the fact that interaction with them is much more social in nature than with other devices found in the home. Pioneering work by Nass et al. showed that a number of phenomena normally associated with human interaction also apply to interactions with computers (e.g. that computers are perceived as social actors), and that voice interfaces only exacerbate this effect, with computer generated voices being perceived as gendered and prompting automatic and unconscious social responses [7, 8].

More recent findings confirm this; similar to interpersonal conversations, interactions with voice assistants can often be positive even when failing to fulfil their functional objectives—the interactions *themselves* are satisfying [9, 10]. Given that Speech activates the same centres in the brain regardless of whether it originates from a home assistant or another person, designing voice controlled systems presents an array of complex functional and ethical challenges.

2.4 Respect and the Socratic Method

By placing it's users and their interests at the centre of the smart home, Aretha represents the first ever instance of a *respectful* smart home device. Drawing on a rich body of philosophical discussion explored in our previous paper on respect in the context of smart devices [11], we believe that the idea of devices respecting their end-users should serve as a strong design goal for smart devices in highly personal settings. While respect has a diverse set of meanings in different contexts, we focus on the characteristics of four formulations of respect that are particularly relevant to smart devices:

- Respecting explicit *directives* (e.g. laws) [12]
- Respecting someone or something as an *obstacle* to one's own goals (e.g. a sailor's respect for the sea) [12]
- Respect as *recognition* of someone's characteristics (e.g. their skill as a painter) [13]
- Respect as an act of *care* (e.g. long term consideration of the welfare of another) [14]

[1] `https://mycroft.ai`
[2] `https://jasperproject.github.io`
[3] `https://www.home-assistant.io`

These categories of respect also provide us with a way to measure progressions of respect, moving from minimalistic adherence to laws and regulations towards deeper and more considerate interactions with users. In addition to this theoretical background, preliminary work on respectful voice assistants has identified a number of benefits that such devices could provide to users [15, 16].

In addition to being grounded in the concept of respect, the conversational models for Aretha use the Socratic method as a framework for engaging with users about their privacy and security preferences. Favoured by Socrates and Plato, the truth of an initial position or thesis is systematically tested with questions intended to challenge assumptions, question reasoning, and fully explore the consequences of different lines of thought. A common approach in therapy (e.g. [17]), Socratic questioning can be used to guide the interlocutor towards a more well-reasoned position. In the case of Aretha, conversations focus on developing the user's understanding of their own privacy preferences, as well as helping them to determine how to best effect them.

2.5 Decentralised Web

This range of wicked' problems with smart home security and privacy have led to speculative works based on a more decentralised architecture for the World Wide Web, such as Databox and Solid. Databox provides users with a container for users to store personal data, which resides on their local network instead of a remote server [18]. This requires platforms to negotiate access to user data instead of storing it themselves. The Solid framework[4] provides a similar model, with applications accessing a user controlled 'POD' which can be hosted anywhere on the web. In line with its philosophical roots, Aretha follows a similar model, ensuring that data does not leave the home without good reason and user consent.

3 Aretha in Detail

Aretha builds on IoT Refine[5], a previous project that acts as a network disaggregator for smart home devices. Itself extending previous work tracing the data sent by smartphone apps [4], IoT Refine clearly shows which devices in the home are sending data to specific companies (see Figure 1). The software operates a WiFi hotspot to which other devices can be connected, performing analysis on this traffic to reveal the companies and countries that data is flowing to, highlighting major changes over time.

By converting this information from the language and concepts of computer networking (e.g. 10 packets to 64.233.160.0) to those that users can intuitively reason about (e.g. 2MB was sent to DoubleClick, a known internet tracker), IoT Refine supports users in developing a situational awareness of their

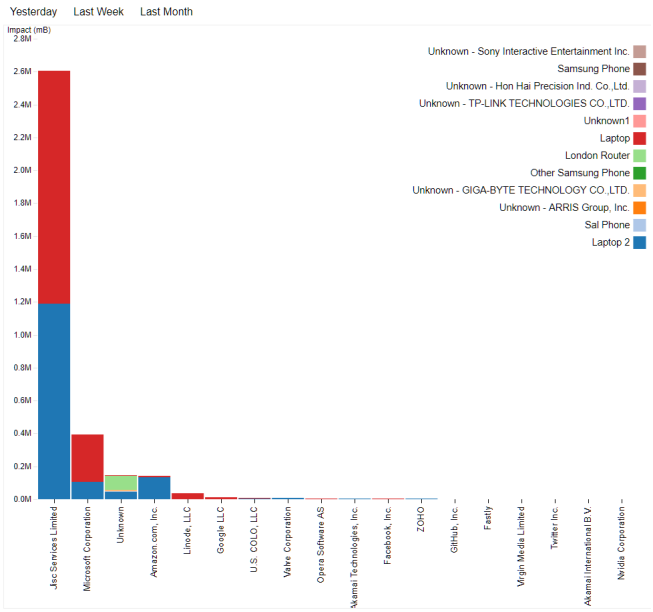

Fig. 1: Traffic Flows in IoT Refine, sorted by company

home network that can then be factored into future decision making, including disconnecting compromised devices.

The following sections describe how Aretha uses spoken conversations generated from IoT Refine data to scaffold user understanding and aid the formation and enactment of personal privacy and security preferences.

3.1 Eliciting Preferences: Establishing a Baseline

Voice interfaces present the perfect opportunity to match the high level information provided by IoT Refine's data flow analysis with the human-readable firewall options described above. Acknowledging privacy and security preferences as fluctuating and idiosyncratic, Aretha approaches the formation and revision of these preferences as an ongoing process, facilitating periodic change rather than treating them as a concrete 'truth'.

The value of information presented by Aretha is diminished if users have a reduced capability to act on it, and similarly, the presentation of options would be of little use if users did not possess the knowledge required to exercise them. The conversation module of Aretha is therefore designed to pair education and exploration with the capacity to act, through the presentation of facts about the user's network with questions designed to help users form and reform their preferences. Where Aretha identifies discrepancies between preferences and device behaviour, users are presented with a suite of options that can be used to address the problem.

To begin with, Aretha attempts to take a snapshot of the user's current preferences in order to establish a baseline level of understanding and concern. This is achieved by asking general questions about the ideal state of their smart home setup, with

[4] `solid.inrupt.com`
[5] `https://github.com/OxfordHCC/IoT-refine`

answers to these questions forming a central thesis around which subsequent questioning is based. A conversation between the **U**ser and **A**retha might begin:

> **A**: How would you feel if your smart devices sold your data to advertisers?
> **U**: I think I'd be fine with it, it's pretty normal, right?

3.2 Questioning Assumptions: Challenging Folk Theories

The questioning phase of the cycle involves Aretha showing examples and counterexamples based on real network traffic from the user's network. This is achieved using Socratic questioning, the goal of which is to guide users in their discovery of their personal preferences by challenging the perceptions and assumptions set out in the elicitation phase described above, rather than to indoctrinate them with a canonical 'right' answer. Thus, the above conversation might proceed as follows:

> **A**: Do you think your current devices send data to advertisers?
> **U**: Yeah, probably
> **A**: What about your Nokia Body scales, do you think they send health data, like your weight, to advertisers?
> **U**: I guess?
> **A**: Is this fine too?
> **U**: Hmm, I think I'd rather it didn't do that

This grounds the abstract concepts from the questioning phase in real data, and helps provide motivation for the user to act if they see fit (in the example above, the user might simply decide that data sharing by the scales is fine, at which point the conversation ends). Through consideration of these potentially contradictory positions (e.g. that the user is concerned about tracking, but one of their devices sends vast amounts of information to known trackers), users are incentivised to use Aretha's network management capabilities to help move the reality of their home network closer to its ideal state.

3.3 Presenting Options: Offering a Way out of the Maze

Finally, Aretha offers users options that correspond to discrepancies identified during the questioning phase. This is vital in utilising the call to action generated by the Socratic method, and helps to reinforce the idea that the user is in control of the situation. Leveraging its position as an IoT hub through IoT Refine, Aretha is able to (a) block traffic to and from specified companies, countries, and industry sectors; (b) block traffic to and from hosts that are known threats; (c) identify when devices are sending traffic unencrypted; and (d) identify when a company that a device has interacted with suffers from a data breach.

In which case the conversation above might continue:

> ...
> **U**: Hmm, I think I'd rather it didn't do that
> **A**: OK, Would you like me to block advertising traffic to and from the Nokia Body scales?
> **U**: Yes please
> **A**: Will do, let me know if this causes any problems

When an action involves blocking traffic to and/or from a device, Aretha will prompt the user in the future to review this decision. This is important as some devices may lose functionality when traffic to certain destinations is blocked.

4 Aretha System Architecture

The Aretha software stack is written in python and nodejs, running Debian 9 (Stretch). The reference hardware platform for deployment during experiments is an Intel NUC small form-factor computer containing an i5-7260U, 16GB DDR4 memory, and a 256GB M2 solid state hard drive.

Designed as an extension to existing voice assistants, Aretha is designed to make use of the speech to text (STT) and text to speech (TTS) capabilities of the host application. Unfortunately, the current status of open source STT and TTS applications, such as Mozilla's DeepSpeech project (used by Mycroft), is still largely experimental and under active development. For this reason, the Aretha prototype is implemented as an Alexa skill. Analysis from the data flow module is accessed via the IoT Refine API, and then a decision tree is used to determine the structure of the next conversation. Planned future work includes the creation of software adaptors from Aretha to stable open voice assistants.

When the outcome of a conversation with Aretha involves one or more actions, these are communicated back to IoT Refine. After performing a reverse lookup of the requested rule (e.g. 'Doubleclick' back to 64.233.160.0), this is entered into iptables with a DROP verdict for future enforcement. A copy of the high and low level versions of any rules the user has chosen to generate are stored so that they can be easily queried or removed in the future. If Aretha discovers a device connecting to a new IP address that matches a company or country already blocked by Aretha, then that new IP address is entered into iptables as above and associated with the original rule in Aretha's database.

5 Evaluation and Future Work

The evaluation of Aretha includes installation at BRE, Watford as part of the PETRAS *IoT in the Home* demonstrator, as well as a planned user study where Aretha will be deployed into a small number of households. The deployment will collect empirical data about how and when Aretha was used, as well as the subjective experiences of participants through interviews and diaries.

In future work we intend to extend the respectful behaviour of Aretha in order to further help users navigate the complex

idiosyncrasies present in the home environment. By exploring how users interact with a highly personal and configurable voice assistant, we hope to move further towards our goal of developing truly respectful devices. Two major planned extensions to Aretha are listed below.

5.1 Decision-making personas

When asking someone to perform a task for us, we have various express and implied expectations about they will perform that task. When making vocal requests to assistants to purchase items or arrange appointments, we make apply social rules and incorrectly make similar assumptions (see voice interface context above). One can imagine a corporate assistant optimising for profit by choosing items with the highest markup, or a more frugal one aggressively lowering the heating during winter to save money. Future versions of Aretha that contain basic searching and purchasing functionality will state upfront what is being optimised for when making decisions, allowing users to mould and shape their Aretha over time to suit their needs.

5.2 Contextual Awareness

The ability of smart devices to observe users across privacy contexts in the home requires them to make many judgements concerning access to information. This includes access to device logs (is it possible to audit a device without infringing the privacy of other users?), as well as content restrictions for groups such as children and teenagers. Future versions of Aretha will present an accessible and configurable model by engaging users in a dialogue about how they want their device to behave. It will also presents a medium through which members of the household can negotiate this behaviour, rather than privileging one user above all others (as is typically the case with account owner).

6 Conclusion

In presenting the design outline for Aretha, we hope to challenge perceptions of what voice assistants can be, and whose interests they can serve. In doing so we contribute a counter narrative to conceptions of voice assistants as seen in the media and scholarly literature, typically as annoyances, liabilities, and tools for surveillance.

The strong philosophical and ethical grounding that underpins Aretha, focused through the vocabulary of respect, offers a chance to break out of our conceptions of the probable trajectory that current voice assistants will take, and move towards a more speculative imagining of what might be plausible in slightly different contexts. From this new beginning, we hope to demonstrate how many of the privacy and security problems present in current voice assistants might be mitigated or precluded entirely.

The conversations that Aretha is able to have with users is not the be all and end all when it comes to discussions about privacy and security. We hope that, in time, the conversations that Aretha has with its users will empower them to share similar experiences with their friends and family, bootstrapping an important contemplative and reflective process in the future.

7 Acknowledgements

This work was supported by the PETRAS IoT Hub Strategic Fund through *ReTiPS: Respectful Things in Private Spaces* and the *IoT in the Home* demonstrator. The PETRAS IoT Hub Strategic Fund is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number N02334X/1.

References

[1] A. Acquisti and J. Grossklags, "Privacy and rationality in individual decision making," *IEEE security & privacy*, vol. 3, no. 1, pp. 26–33, 2005.

[2] P. A. Norberg, D. R. Horne, and D. A. Horne, "The privacy paradox: Personal information disclosure intentions versus behaviors," *Journal of Consumer Affairs*, vol. 41, no. 1, pp. 100–126, 2007.

[3] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir, and H. Borgthorsson, "Leakiness and creepiness in app space: Perceptions of privacy and mobile app use," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2347–2356, ACM, 2014.

[4] M. Van Kleek, R. Binns, J. Zhao, A. Slack, S. Lee, D. Ottewell, and N. Shadbolt, "X-ray refine: Supporting the exploration and refinement of information exposure resulting from smartphone apps," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 393, ACM, 2018.

[5] F. Brunton and H. Nissenbaum, *Obfuscation: A user's guide for privacy and protest*. Mit Press, 2015.

[6] H. Nissenbaum, *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[7] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," in *Proceedings of the CHI conference on Human factors in computing systems*, CHI '94, pp. 72–78, ACM, 1994.

[8] B. Reeves and C. I. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, 1996.

[9] A. Purington, J. G. Taft, S. Sannon, N. N. Bazarova, and S. H. Taylor, "Alexa is my new bff: Social roles, user satisfaction, and personification of the amazon echo," in *Proceedings of the 2017 CHI Conference Extended*

Abstracts on Human Factors in Computing Systems, CHI '17, pp. 2853–2859, ACM, 2017.

[10] I. Lopatovska, K. Rink, I. Knight, K. Raines, K. Cosenza, H. Williams, P. Sorsche, D. Hirsch, Q. Li, and A. Martinez, "Talk to me: Exploring user interactions with the amazon alexa," *Journal of Librarianship and Information Science*, 2018.

[11] M. Van Kleek, W. Seymour, R. Binns, and N. Shadbolt, "Respectful things: Adding social intelligence to 'smart' devices," in *Living in the Internet of Things: Cybersecurity of the IoT*, IET, 2018.

[12] S. D. Hudson, "The nature of respect," *Social Theory and Practice*, vol. 6, no. 1, pp. 69–90, 1980.

[13] S. L. Darwall, "Two kinds of respect," *Ethics*, vol. 88, no. 1, pp. 36–49, 1977.

[14] R. S. Dillon, "Respect and care: Toward moral integration," *Canadian Journal of Philosophy*, vol. 22, no. 1, pp. 105–131, 1992.

[15] W. Seymour, "How loyal is your alexa?: Imagining a respectful smart assistant," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, 2018.

[16] W. Seymour, "Privacy therapy with aretha: What if your firewall could talk?," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, 2019.

[17] C. A. Padesky, "Socratic questioning: Changing minds or guiding discovery," in *A keynote address delivered at the European Congress of Behavioural and Cognitive Therapies, London*, vol. 24, 1993.

[18] A. Crabtree, T. Lodge, J. Colley, C. Greenhalgh, R. Mortier, and H. Haddadi, "Enabling the new economic actor: data protection, the digital economy, and the databox," *Personal and Ubiquitous Computing*, vol. 20, no. 6, pp. 947–957, 2016.