control. By allowing users to negotiate alternative data, giving them more control over their personal information. As such, users can provide a less revealing version of the requested data, *e.g.* limiting sensitive information while still accessing services.

```
+!pr(dob, vibe)[perform]:{True} ←                          1
  +substitute(dob, age_range),                             2
  send(user, :perform, substitute(dob, age_range));        3
```

Listing 5: Data collection substitution plan

Home Security Risks. While both home address and GPS data pose privacy risks individually, the risk to privacy arises significantly when home address data is used in combination with GPS data. The home address reveals 'where' you live, but GPS data can reveal 'when' you are away. When combined, these pieces of information allow for more precise targeting of crimes such as burglaries. To prevent it, these two data can be disassociated given in Listing 6.

```
+!pr(GPS, vibe)[perform]:{                                 1
 B request(home_address, vibe)} ←                          2
 +decline(GPS),.send(user, :perform, decline(gps));        3
```

Listing 6: Data collection decline plan for GPS

Health Insurance Discrimination. By analysing the user's daily physical activities, workout plans, dietary recommendations, date of birth, gender, and BMI, the app could infer potential health risks or medical conditions. This sensitive information could be shared with insurance companies, which could result in higher premiums or even denial of coverage. A quick way to avoid this is to disassociate the data of bmi from the rest of these data shown in Listing 7.

```
+!pr(bmi, vibe)[perform]:{                                 1
 B request(dob, vibe), B request(gender, vibe),            2
 B workout_plan(present),                                  3
 B diet_recommendation(present),                           4
 B daily_physical_activities(monitored) } ←                5
 +decline(bmi),.send(user, :perform, decline(bmi));        6
```

Listing 7: Data collection decline plan for BMI

Finally, we have a plan in Listing 8 to approve other personal data that are not already subject to any specified action control. In Gwendolen, plans listed first are given higher priority. The plan in Listing 8, which is put at the list's end, will only execute if no preceding plans, such as data declination, apply.

```
+!pr(D,N)[perform]:{True} ←                                1
   +approve(D),.send(user, :perform, approve(D));          2
```

Listing 8: Data collection approve plan

Verification Capability. We have verified standard safety and liveness properties. For example, if the set of action controls available includes "approve, decline, substitute" data requests. Then, the safety properties ensure that the privacy agent will neither approve, decline, nor substitute a data request all at once (safety) and that if data were requested, there would always be some response eventually (liveness). Some of these formalisation are given as follows where _ denotes a wildcard which is supported natively by MCAPL:

safety property:

$$[](\sim B(ag, approve(gps)) \;||\; \sim B(ag, decline(gps)) \;||\; \sim B(ag, substitue(gps, \_)))$$

liveness property:

$$[](B(ag, request(gps, vibe)) \rightarrow <>(B(ag, approve(gps)) \;||\; B(ag, decline(gps)) \;||\; B(ag, substitute(gps, \_))))$$

We also proved properties specific to the user's privacy requirements. For example, if the privacy agent believes that the data of home address B(ag, request(gps, vibe)) and GPS are under request by the app *vibe* and the agent has not either approved GPS or substituted it with something else, the data of GPS will be declined eventually. This property can be formalised as follows:

$$[](B(ag, request(home\_address, vibe)) \;\&\; B(ag, request(gps, vibe)) \;\&\; \sim B(ag, approve(gps)) \;\&\; \sim B(ag, substitute(gps, \_)) \rightarrow <>\; B(ag, decline(gps)))$$

The implementation and the full list of proven properties can be found in the GitHub distribution[4].

6 EVALUATION

We provide an in-depth evaluation of our approach both 1) in simulations to demonstrate its capability for run-time privacy decision-making and 2) in verification to show the computational complexity to guarantee the correctness of the design of such agents. Here by *simulation*, we mean simply running an agent and *verification* means analysing an agent. We are particularly interested in simultaneous data requests for simulation evaluation to ensure our approach can indeed relieve humans with this overwhelming data request in an acceptable time. Meanwhile, we shed insights into how expensive verification can be and how the computational expensiveness varies in different settings. All results are obtained on a laptop with a 16-core Intel Core i7-11800H at 2.30GHz (hyperthreaded), 16 GB memory, and running 64-bit Ubuntu Linux 20.04.3 LTS. Again, the full source codes can be found in the previous GitHub link.

6.1 Evaluation Setting

In the interests of generality, our evaluation is based on a set of randomly generated, synthetic parameterised data requests and relevant Gwendolen plans representing the privacy requirements to address these data . By varying the number of data which may be requested and plans for these data requests, and whether data is requested in a simultaneous (*i.e.* data requested at the same time) or sequential (*i.e.* data requested one after the other) manner, we can evaluate the performance of our approach. For simplicity, we assume an environment that only accounts for data requests.

6.2 Simulation

To cope with the high volume of data a user may have to consider giving consent to share in an acceptable time, we *stress test* out our approach for handling data requests in a simultaneous manner. The results are shown in Fig. 2. The left plot in Figure 2 which varies the number of data requests but assumes one plan for each data request shows its near-linear[5] performance regarding the number of data requests. For example, it took 3 seconds to respond to 100 data but around 4 minutes to respond to 5000 data. Meanwhile, the

[4] https://github.com/Mengwei-Xu/privacy_agents_mcapl

[5] The coefficient of determination for the linear regression analysis is around 0.918.

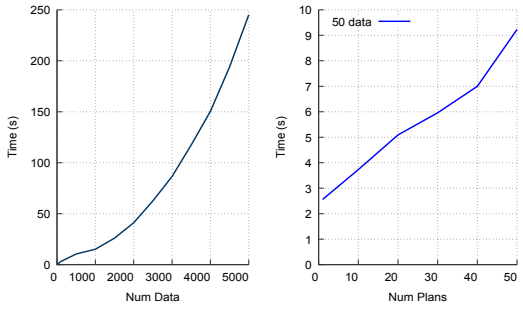

Figure 2: Simulation time increases near-linearly with the number of data and plans for simultaneous data requests.

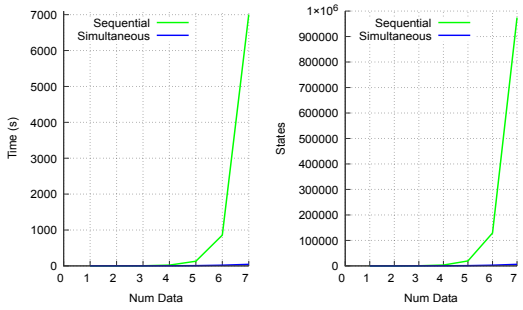

Figure 3: Transition system construction time and states increase exponentially with the number of data.

right plot in Figure 2 shows the near-linear increase of time when the number of plans increases in the case of 50 data requests. This is because the agent may need to search each plan one by one to find an applicable plan to address a given data request.

6.3 Verification

Unlike the simulation, verification is highly expensive due to its exhaustive nature in general. A crucial contribution of our evaluation is to be able to provide insights on the contrasting levels of difficulties depending on how data is requested and what level of confidence we are seeking. For example, Fig. 3 shows that it is significantly more expensive to verify our privacy agent in the sequential setting than the simultaneous one. It should come unsurprised as in the sequential data requests, we are verifying all possible subsets of data that have not been requested at all possible steps. On the contrary, in the simultaneous data requests, the complexity mainly comes at the beginning where all possible subsets of data may be requested. However, though significantly cheaper to verify than the sequential data requests, the simultaneous data requests are still exponential in their nature shown in Fig. 4.

7 DISCUSSION AND CONCLUSION

Our approach casts data collection at the user's end as a decision-making problem and employs autonomous agents to bear this responsibility. While this delegation may enhance efficiency, the trustworthiness of these agents becomes a crucial concern. Therefore,

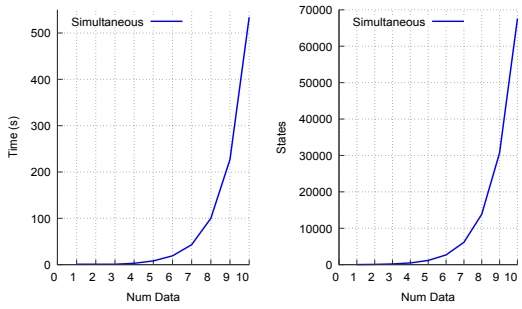

Figure 4: Transition system construction time and states increase exponentially with the number of data.

formal verification becomes indispensable to establish the confidence of these autonomous agents. In contrast to most of the previous work, we can guarantee the correct agent behaviours to handle highly sensitive personal data at the stage of data collection.

Nevertheless, our approach comes with some limitations. We note that the agent verification in the sequential setting will quickly and inevitably cause a state explosion. In practice, since verification is done at the design time, it will not cause any issues in the actual operation (shown in the simulation). That said, this issue can be mitigated by focusing on a small set of environmental inputs that will affect one another. Another issue we have not considered is the potentially conflicting privacy requirements. To address this, we have some ideas to adopt a priority-based conflict-solution framework. If it is bound to violate two privacy requirements, the agent should violate the one which will cause less damage. And Gwendolen naturally supports order-based plan selection. Finally, our framework does face another inevitable privacy threat: the possibility of an unauthorised intrusion into our privacy agent, potentially exposing user privacy requirements. Employing encryption protocols to mitigate this is a future consideration.

Overall, our approach is abstract and generic in nature, supporting both simultaneous and sequential data requests, and we have supplied a computational instantiation of the framework for proof-of-concept through autonomous agents together with some in-depth evaluation of its run-time decision practicality and expensive verification complexity. The use of autonomous systems not only highlights their capacity to tackle societal issues but also emphasises the need for user trust. Our approach showcases the intuitive alignment and effectiveness of a symbolic cognitive agent (which we can trust through strong proof) in addressing the typical data collection challenges users encounter daily.

ACKNOWLEDGMENTS

This work is supported by the University of Manchester and Newcastle University, and EPSRC, under projects Computational Agent Responsibility (EP/W01081X), UKRI Trustworthy Autonomous Systems Node in Verifiability (EP/V026801/2), and EnnCore: End-to-End Conceptual Guarding of Neural Architectures (EP/T026995/1).

REFERENCES

[1] Rakesh Agrawal and Ramakrishnan Srikant. 2000. Privacy-preserving data mining. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 439–450.

[2] Khaled Gubran Al-Hashedi and Pritheega Magalingam. 2021. Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review* 40 (2021), 100402.

[3] Rob Ashmore, Radu Calinescu, and Colin Paterson. 2021. Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *ACM Computing Surveys (CSUR)* 54, 5 (2021), 1–39.

[4] Masoud Barati, Gagangeet Singh Aujla, Jose Tomas Llanos, Kwabena Adu Duodu, Omer F Rana, Madeline Carr, and Rajiv Ranjan. 2021. Privacy-aware cloud auditing for GDPR compliance verification in online healthcare. *IEEE Transactions on Industrial Informatics* 18, 7 (2021), 4808–4819.

[5] Lemi Baruh, Ekin Secinti, and Zeynep Cemalcilar. 2017. Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication* 67, 1 (2017), 26–53.

[6] Rajdeep Bhanot and Rahul Hans. 2015. A review and comparative analysis of various encryption algorithms. *International Journal of Security and Its Applications* 9, 4 (2015), 289–306.

[7] Michael Bratman. 1987. Intention, plans, and practical reason. (1987).

[8] Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The guardian* 17, 1 (2018), 22.

[9] Nick Couldry and Ulises A Mejias. 2019. The costs of connection. In *The Costs of Connection*. Stanford University Press.

[10] Ludivine Crépin, Yves Demazeau, Olivier Boissier, and François Jacquenet. 2009. Sensitive data transaction in hippocratic multi-agent systems. In *Engineering Societies in the Agents World IX: 9th International Workshop*. Springer, 85–101.

[11] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. 2018. Personalized privacy assistants for the internet of things: Providing users with notice and choice. *IEEE Pervasive Computing* 17, 3 (2018), 35–46.

[12] Louise A Dennis. 2017. Gwendolen semantics: 2017. (2017).

[13] Louise A Dennis. 2018. The MCAPL Framework including the Agent Infrastructure Layer and Agent Java Pathfinder. *The Journal of Open Source Software* 3, 24 (2018).

[14] Louise A Dennis, Michael Fisher, Matthew P Webster, and Rafael H Bordini. 2012. Model checking agent programming languages. *Automated software engineering* 19 (2012), 5–63.

[15] E Allen Emerson. 1990. Temporal and modal logic. In *Formal Models and Semantics*. Elsevier, 995–1072.

[16] Lujun Fang and Kristen LeFevre. 2010. Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*. 351–360.

[17] Michael R. Galbreth and Mikhael Shor. 2010. The Impact of Malicious Agents on the Enterprise Software Industry. *MIS Quarterly* 34, 3 (2010), 595–612.

[18] Michael P Georgeff and A Rao. 1992. An abstract architecture for rational agents. In *Proc. of the Third International Conference on Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers Inc San Francisco, CA, USA, 439–449.

[19] Mohamad Gharib, Paolo Giorgini, and John Mylopoulos. 2020. An ontology for privacy requirements via a systematic literature review. *Journal on Data Semantics* 9 (2020), 123–149.

[20] Osman Hasan and Sofiene Tahar. 2015. Formal verification methods. In *Encyclopedia of Information Science and Technology, Third Edition*. IGI Global, 7162–7170.

[21] Özgür Kafalı, Akın Günay, and Pınar Yolum. 2014. Detecting and predicting privacy violations in online social networks. *Distributed and Parallel Databases* 32, 1 (2014), 161–190.

[22] Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.

[23] Nadin Kökciyan, Nefise Yaglikci, and Pinar Yolum. 2017. An argumentation approach for resolving privacy disputes in online social networks. *ACM Transactions on Internet Technology (TOIT)* 17, 3 (2017), 1–22.

[24] Nadin Kökciyan, Pınar Yolum, et al. 2022. Taking situation-based privacy decisions: Privacy assistants working with humans. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*. 703–709.

[25] A Can Kurtan and Pınar Yolum. 2021. Assisting humans in privacy management: an agent-based approach. *Autonomous Agents and Multi-Agent Systems* 35 (2021), 1–33.

[26] Pierre M Nugues. 2006. *An introduction to prolog*. Springer.

[27] P Blessed Prince and SP Jeno Lovesum. 2020. Privacy enforced access control model for secured data handling in cloud-based pervasive health care system. *SN Computer Science* 1, 5 (2020), 239.

[28] Anand S Rao. 2005. AgentSpeak (L): BDI agents speak out in a logical computable language. In *7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*. Springer, 42–55.

[29] Anand S Rao and Michael P Georgeff. 1997. Modeling rational agents within a BDI-architecture. *Readings in agents* (1997), 317–328.

[30] Anand S Rao and Michael Wooldridge. 1999. *Foundations of rational agency*. Springer.

[31] Protection Regulation. 2018. General data protection regulation. *Intouch* 25 (2018), 1–5.

[32] Ralph Schäfermeier, Theodoros Mitsikas, and Adrian Paschke. 2022. Modeling a GDPR Compliant Data Wallet Application in Prova and AspectOWL. *on AI Compliance Mechanism (WAICOM 2022)* (2022), 50.

[33] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. 2020. Circumvention by design-dark patterns in cookie consent for online news outlets. In *Proceedings of the 11th nordic conference on human-computer interaction: Shaping experiences, shaping society*. 1–12.

[34] Monica Tentori, Jesus Favela, and Marcela D Rodriguez. 2006. Privacy-aware autonomous agents for pervasive healthcare. *IEEE Intelligent Systems* 21, 6 (2006), 55–62.

[35] Onuralp Ulusoy and Pinar Yolum. 2021. Panola: A personal assistant for supporting users in preserving privacy. *ACM Transactions on Internet Technology (TOIT)* 22, 1 (2021), 1–32.

[36] Jaideep Vaidya and Chris Clifton. 2004. Privacy-preserving data mining: Why, how, and when. *IEEE Security & Privacy* 2, 6 (2004), 19–27.

[37] Samuel Warren and Louis Brandeis. 1989. The right to privacy. In *Killing the Messenger*. Columbia University Press, 1–21.

[38] Alan Westin. 1984. The origins of modern claims to privacy. (1984).

[39] Alan F Westin. 1968. Privacy and freedom. *Washington and Lee Law Review* 25, 1 (1968), 166.

[40] Michael Wooldridge. 2009. *An introduction to multiagent systems*. John wiley & sons.

[41] Ni Zhang and Chris Todd. 2006. A privacy agent in context-aware ubiquitous computing environments. In *10th IFIP TC-6 TC-11 International Conference*. Springer, 196–205.