Analysis of Data Disclosure in Asynchronous Distributed Systems

Fatemeh Ghassemi

University of Tehran

Abstract: People provide their information to distributed systems to receive the desired services. This information may be disclosed to the agents of the system as part of messages transmitted among them. As the agents of the system are smart, they can infer new information from their obtained information, that they may not be authorized to know. So, preserving privacy in such systems is an important yet challenging issue. We study the problem of analyzing the disclosure of private information in distributed asynchronous systems. We assume the system follows a set of policies defined at design time to prevent private information disclosure.

To achieve this, we construct a model of the system, consider the policies as the system properties, and check whether these properties are satisfied in the system using the model checking technique. To construct a model of the system, we extend the actor model, which is a well-known reference model for distributed asynchronous systems, by enriching actors with the knowledge base and inference capability.

Keywords: Distributed systems, Knowledge, Inference, Privacy Policy, Formal Verification

1 Introduction

Today, the use of distributed systems, as in Internet of Things (IoT) or microservice architectures, is growing and people provide their information to these systems to receive the desired services. In these systems, personal information is collected by various processes and devices. The information may be shared with many service providers to be analyzed or reported for further objectives [1]. The information may also be provided to other parties for various intentions such as research or marketing. In such systems, if there is no sufficient control over the transmitted data among agents, a personal data breach may happen, so preserving privacy in such systems is an important and yet challenging issue. For example, a patient provides her personal information to a health care system for treatment, but if this information is sent to a research organization without the patient's consent, a personal data breach has occurred.

Privacy violations can happen in different ways. Information collection, information processing, information dissemination, and invasions are different types of privacy violations [2]. Disclosure is a special form of information dissemination, which means ``making private information known outside the group of individuals expected to know it'' [3]. A useful method to prevent private information disclosure would be to define policies which control the disclosure of this information in the system and require the system to follow those policies.

As stated in [4], if the policies for protecting the privacy of individuals are violated, it is not only harmful to the individuals whose information is being disclosed, but can also be damaging to the organization that violates these policies. Therefore, having a framework with formal foundation to ensure that the system works according to its defined policies, is valuable.

Privacy is correlated with the interaction aspects of distributed systems [1]. Disclosure of personal information occurs when an agent receives information. The ways in which an agent can receive information about other agents can be classified into three categories: direct receive, indirect receive, or receive by inference [5,6]. The difference between direct and indirect receives is that in the first case, the owner of personal information directly sends its information to another agent, but in the second case an agent sends the personal information of another agent to a third one. In \textit{receive by inference}, the agent infers other agents' personal information based on the information received previously from other agents. We introduce an approach that checks the disclosure of sensitive information as the result of inference, as well as direct and indirect receive, in the distributed systems.

2 Research Goals

With the aim to check the disclosure of sensitive information as the result of inference, as well as direct and indirect receive, in the distributed systems, we are going to use model checking to analyze information disclosure, i.e., policies, in the presence of inference capability for the agents in the domain of distributed systems.

To make the model checking approach feasible, we need a modeling notation that is suitable for specifying and analyzing such systems. We base our modeling approach on *Actor model* [7], which is a well-known computation model for concurrent and distributed systems. An actor model consists of a set of active objects called *actors*, which encapsulate data, communicate via asynchronous message passing, and have no shared data. These characteristics are naturally suitable for modeling distributed systems in the real world. The actor model guarantees delivery of the messages, but the order in which the actors execute and the order of receiving messages, which are sent by different actors to a specific actor, are nondeterministic. This nondeterminism models the delays and effects of the network in sending messages. There are a number of actor-based programming and modeling languages like Erlang [8], Rebeca [9], Ptolemy II [10], and ABS [11] proposed for different design concerns. For example, Rebeca and ABS are designed for analysis and code generation [12], while Erlang is optimized for efficient execution. However, modeling and analysis of privacy and information disclosure have not been the design concern of any of them. In the actor model, the actors' information can be disclosed among other actors as part of the transmitted messages, so it is essential to protect actors' private information from disclosure to unauthorized actors. We propose an actor-based modeling language to model the actors' knowledge and inference capability and define policies that enable one to specify restrictions over the actors' knowledge to avoid disclosure of private information to unauthorized actors.

Reference

[1] Samani, A.: “Privacy in Cooperative Distributed Systems: Modeling and Protection Framework”; *Ph.D. dissertation*, The University of Western Ontario (2015).

[2] Solove, D.J.: “A Taxonomy of Privacy”; *University of Pennsylvania Law Review*, Vol. 154, No. 3, (2006), 477–560.

[3] Tschantz, M. and Wing, J.: “Formal Methods for Privacy”; *In Proc. 2nd World Congress on Formal Methods*, Springer-Verlag, Berlin, Heidelberg (2009), 1–15

[4] Ronne, J.: “Leveraging Actors for Privacy Compliance”; *In Proc. 2nd edition on Programming systems, languages and applications based on actors, agents, and decentralized control abstractions*, ACM (2012), 133–136

[5] Riahi, Sh., Khosravi, R., and Ghassemi, F.: “Purpose-based Policy Enforcement in Actor-based Systems”; *In Proc. 7th International Conference on Fundamentals of Software Engineering*, LNCS, Springer (2017), 196–211

[6] Blanke, J.M.: “Protection for ‘Inferences Drawn’: A Comparison Between the General Data Protection Regulation and the California Consumer Privacy Act’’; *Global Privacy Law Review*, Vol. 1, Issue 2, (2020), 81–92

[7] Agha, G.A.: “Actors - a model of concurrent computation in distributed systems”; *MIT Press series in artificial intelligence*, MIT Press (1985).

[8] Armstrong, J.: “*Programming Erlang, Software for Concurrent World*”; Pragmatic Bookshelf (2007).

[9] Sirjani, M., Movaghar, A., Shali, A., and de Boer, F.: “Modeling and verification of reactive systems using Rebeca”; *Fundamenta Informaticae* 63, (2004), 385–410.

[10] Eker, J., Janneck, J., Lee, E.A., Liu, J., Liu, X., Ludvig, J., Neuendorffer, S., Sachs, S., and Xiong, Y.: “Taming heterogeneity - the Ptolemy approach”; *In Proc. the IEEE, Vol. 91, Issue 1, January*, 127–144

[11] Abstract Behavioral Specification (ABS) language, http://abs-models.org.

[12] Boer, F.D., Serbanescu, V., Hähnle, R., Henrio, L., Rochas, J., Din, C.C., Johnsen, E.B., Sirjani, M., Khamespanah, E., Fernandez-Reyes, K., and Yang, A.M.: “A survey of active object languages”; *ACM Computing Surveys (CSUR)*, Vol. 50, No. 5, (2017), 1–39.