Agent-based Epistemic Secrecy

Patrick Krümpelmann and **Gabriele Kern-Isberner**
Information Engineering Group, Technische Universität Dortmund, Germany

Abstract

Secrecy in multiagent systems is often defined from an abstract, global perspective with unrealistic presumptions. The resulting notions of secrecy are very strict and heavily constrain the information flow. Here, we approach the topic of secrecy from the point of view of an autonomous epistemic agent with incomplete and uncertain information. It is situated in a multiagent system, needs to communicate with other agents to achieve its goals, but has agent-specific secrets with varying strength it does not want to reveal. We develop a framework for secrecy based on the local epistemic state of an agent and define an agent-based notion of secrecy. We relate our notion of secrecy to other approaches to secrecy, formally show the relationship, and discuss the advantages of our approach.

Introduction

A large body of work exists on the topic of secrecy and diverse definitions of secrecy in various settings with different properties have been developed. Hereby it can be noted that research in secrecy is focused on strong notions of secrecy of a whole (multiagent) system. Secrecy is defined on a global, static, and complete view of the system and hardly permits any information flow. Agents are, if at all, considered as simple entities. Realistic scenarios of autonomous intelligent agents in dynamic, uncertain environments, however, do not meet the prerequisites for general, global definitions of secrecy. The latter are too strict for realistic scenarios. As observed in (Halpern and O'Neill 2008) a major task for future work on secrecy is the "careful consideration of how secrecy definitions can be weakened to make them more useful in practice". In this work we consider secrecy from the point of view of an autonomous epistemic agent with incomplete and uncertain information which is situated in a multiagent system. The agent pursues its goals by performing actions in its environment which naturally includes communication with other agents. On one hand, the exchange of information with other agents is often essential for an agent in order to achieve its goals. Especially if the agent is part of a coalition. On the other hand the agent is interested, or obliged, not to reveal certain information, its secrets. Restriction of communication leads to a loss in performance and utility of the individual agents, coalitions and the whole multiagent system. Therefore a good solution of the implied conflict between the agent's goal to preserve secrecy and its other goals is one that restricts communication as little as necessary in order to preserve secrecy.

In realistic settings the information to be kept secret is neither global, i. e. uniform, nor static. Secrets are not global in their content as an agent has different secrets with respect to different agents. They are also not global with respect to their strength. That is, an agent wants to keep some information more secret as other. These differences in strength of secrets arise naturally from the value of the secret information. The value of secret information depends on the severeness of the negative effects, or the cost, for the agent resulting from disclosure of the secret information. These costs can differ widely and consequently the agent is interested in not revealing secret information to different degrees. Secrets are also not static, they arise, change and disappear during runtime of an agent such that it has to be able to handle these changes adequately. Apart from being aware of its secrets at any time, an agent has to act such that it avoids revealing its secrets. It shall only reveal secret information if it considers it necessary in order to achieve its goals; which depends on the strength of the secret and the utility of the goals. That is, an agent has to take its secrets into account while acting and, moreover, while planing its intended course of action.

The main contribution of this paper is the approach of the topic of secrecy from the epistemic agent's perspective. We develop an adequate notion of secrecy for epistemic agents in multiagent systems that satisfies the requirements laid out above. To this end, we develop a general epistemic agent model for secrecy. An agent therein is characterized by an action function which determines its behavior. We base our notion of a secrecy preserving agent on an action function which prevents the disclosure of secrets. Furthermore, we instantiate our framework for the widespread runs-and-systems framework and show that our framework generalizes other notions of secrecy and indeed allows for handling uncertainty in secrecy. We see our work as the exploration of the topic of secrecy from a subjective perspective under uncertainty.

The rest of this paper is organized as follows. In the following section we develop our epistemic agent model and our notion of secrets. Following on this we define agent based secrecy formally. Using the developed framework, we present our view on secrecy in the runs-and-systems frame-

work and proof formal relations to our framework. Subsequent, we discuss how an agent can act to preserve secrecy. In the last two sections we discuss related work and conclude.

Epistemic Secrecy

In this section we present an epistemic characterization of secrecy for autonomous agents in multiagent systems. We base our definition of secrecy on the view of a single agent that wants to maintain secrecy while interacting with other agents and its environment. It has incomplete and uncertain information. Thus it has to rely on its plausible but defeasible beliefs which result from its current epistemic state, and has to handle dynamics of beliefs and secrets. As motivated above, an agent needs a definition of secrets that allows to define secrets individually for each of its fellow agents. Moreover, each secret can vary in strength, that is, the agent wants to keep some information more secret than other. For the illustration of our approach we use the following running example.

Example 1 (Running Example) *Agent Alfred $\mathcal{A}$ is married to Beatriz $\mathcal{B}$ and has two children with her. Since some weeks ago an affair between him and Carla $\mathcal{C}$, a new colleague of him, has been evolving. Alfred has two main concerns, in no case Beatriz shall even be suspicious of him having an affair. He also does not want that Carla knows that he has children since that might put her off.* [1]

Epistemic State

We assume a multiagent system with a set of agents $\mathfrak{A}$. Each agent has a complex epistemic state including views of the epistemic states of other agents in the set of agents $\mathfrak{A}$. We use the agent identifier $\mathcal{X}$ to denote an arbitrary agent. For the representation of the secrecy scenario it is convenient to restrict the view to the communication between two agents, the modeled agent which wants to defend its secrets, denoted by $\mathcal{D}$, from a potentially attacking agent, denoted by $\mathcal{A}$. Hereby we disregard potential other sources of information of $\mathcal{A}$ other than $\mathcal{D}$, which simplifies our explanations and definitions for now and will be addressed in future work. The epistemic state of an agent contains a complex representation of the agent's current belief state which might contain extra logical information. We keep our representation abstract but assume some underlying languages for different parts of the epistemic state.

Definition 1 (Epistemic State) *Let $\mathcal{X} \in \mathfrak{A}$ be some agent. The epistemic state of $\mathcal{X}$ is denoted by $\mathcal{K}_\mathcal{X}$. The set of all epistemic states of the agents in $\mathfrak{A}$ is denoted by Ω. Agent $\mathcal{X}$'s view on the world is given by $f_W(\mathcal{K}_\mathcal{X}) \subseteq \mathcal{L}_V$. The view agent $\mathcal{X}$ has of the epistemic state of agent $\mathcal{Y} \in \mathfrak{A}$ is given by $f_\mathcal{Y}(\mathcal{K}_\mathcal{X}) \subseteq \mathcal{L}_V$. We presuppose a function $\mathcal{S}(\mathcal{K}_\mathcal{X}) \subseteq \mathcal{L}_S$ which returns the secrets of an epistemic state.*

[1] While this soap opera example is ethically questionable, it exposes a variety of general issues of secrecy that are easily understandable without expert knowledge of a particular domain.

An epistemic state is an abstract representation of the subjective beliefs of the agent. In particular these beliefs contain, or entail information about its secrets and views on other agents. We do not fix the underlying language here, but we think that non-classical formalisms are needed for adequate handling of incomplete and uncertain information, such as conditionals (Kern-Isberner and Krümpelmann 2011) or logic programs (Delgrande et al. 2008). Extra logical information might be represented in form of preference information over the belief base, for example in form of epistemic entrenchment relations, cf. (Gärdenfors and Makinson 1988). Since we are in a multiagent setting epistemic states might as well contain further information about fellow agents such as their credibility, cf. (Krümpelmann and Kern-Isberner 2008; Tamargo et al. 2012).

The abstract formulation of an epistemic state of an agent reflects the intuition that we have to reason in order to come up with a definition of our secrets, or to state our view on the beliefs of others. In this sense the functions $\mathcal{S}(\mathcal{K}_\mathcal{D})$ and $f_\mathcal{A}(\mathcal{K}_\mathcal{D})$ represent these reasoning process on a monolithic epistemic state. We consider this a desirable approach but do not go into detail on this here and focus on the definition of secrecy. To this end we can simplify the epistemic state by assuming an explicit representation of it.

Definition 2 (Compound epistemic state) *Let $\mathcal{X} \in \mathfrak{A}$ be an agent in a given set of agents $\mathfrak{A}$. The compound epistemic state of $\mathcal{X}$ has the form:*

$$\mathcal{K}_\mathcal{X} = \langle B, \mathcal{S}, \{\mathcal{V}_\mathcal{Y} \mid \mathcal{Y} \in \mathfrak{A}\}\rangle.$$

B is the belief base, $\mathcal{S}$ the set of secrets and $\mathcal{V}_\mathcal{Y}$ the view on agent $\mathcal{Y}$ of agent $\mathcal{X}$. Accordingly $f_\mathcal{Y}(\mathcal{K}_\mathcal{X}) = \mathcal{V}_\mathcal{Y}$. Let the set of all possible belief bases be given by $\mathbb{B}$ and the set of all possible secrets by $\mathbb{S}$, defined over the language $\mathcal{L}_{BS}$ of the beliefs.

To equip agents with such reasoning facilities, we define a *belief operator* which returns a set of beliefs given some base representation of a view.

Definition 3 (Belief Operator) *A belief operator is a function $Bel_\mathcal{X} : \mathcal{L}_V \to \mathcal{L}_{BS}$ such that for a given view V we get $Bel(V) \subseteq \mathcal{L}_{BS}$.*

In contrast to the classic notion of a belief set we do not demand it to be deductively closed as we especially want to support non-monotonic logics whose models are usually not closed under deduction. To sum up, three types of languages are used, $\mathcal{L}_V$ for the view of the agent on other agents and its environment, $\mathcal{L}_S$ to define secrects and $\mathcal{L}_{BS}$ for the belief set of some view.

The belief operator determines how the uncertainty of information is handled. Hereby the allowed amount of uncertainty of the beliefs can be defined via the choice of a belief operator. For this sake we define a family $\mathcal{B}$ of belief operators and a linear, i. e. transitive, antisymmetric and total, order $\prec_\mathcal{B}$ on it such that for each $Bel_i, Bel_j \in \mathcal{B}$ we have $Bel_i \prec_\mathcal{B} Bel_j$ or $Bel_j \prec_\mathcal{B} Bel_i$. The definition of a family of belief operators abstracts form the underlying formalism and inference mechanism. Thereby it captures a wide range of formalisms from purely qualitative ones to plausibilistic

ones. The order on the operators hereby represents the uncertainty of the drawn conclusions as illustrated in the following example.

Example 2 *A simple and well known family of operators would be $\mathcal{B}_a = \{Bel_{skep}, Bel_{cred}\}$ consisting of a skeptical and a credulous belief operator for some formalism. The order on these operators can be based on the subset relation, i. e. $Bel_i \prec_{\mathcal{B}} Bel_j$ if and only if $Bel_i(V) \subseteq Bel_j(V)$ for all $V \subseteq \mathcal{L}_V$. Assuming that an agent believes credulously everything that it believes skeptically we get $Bel_{skep} \prec_{\mathcal{B}} Bel_{cred}$.*

A richer hierarchy is given by quantitive approaches as by probabilistic formalisms where a family of operators is given by $\mathcal{B}_b = \{Bel_x \mid x \in [0,1]\}$. Each Bel_x includes every sentence which is believed with probability $\geq x$ and $Bel_x \prec_{\mathcal{B}} Bel_y$ if and only if $x \geq y$.

Secrets

We consider secrets to be highly subjective. From the point of view of an agent it is natural that it knows about its secrets which are generally not global but local, i. e. the agent does not want to share some information with some specific agents.

Example 3 *In our example, Carla is the only person Alfred does not want to know that he has children while nobody apart from her shall know about their affair.*

Of course, agents should take into consideration that other agents may talk to each other, so that secrets may be revealed not only by direct communication. In order to keep our example simple we ignore this here but may well extend our considerations to apply our methods for preserving secrecy to include such communications.

As discussed before, agents may consider some secrets more confidential than others. We reflect this by assigning a specific belief operator from a family of operators to each secret by which it should not be inferable. Hereby, the use of a more credulous belief operator leads to stronger protection of the corresponding secret.

Example 4 *Alfred does not want Beatriz to be suspicious while he considers it sufficient that Carla does not know for sure that he has children. Also he wants to keep his affair secret with varying strength depending on the other agent's relation to Beatriz and does not care what complete strangers think.*

We formalize secrets as triples specifying the information to be kept secret, the belief operator to use and the agent towards which the agent holds the secret.

Definition 4 (Secrets) *A* secret *is a tuple $(\Phi, Bel, \mathcal{A}')$ which consists of a formula Φ, a belief operator Bel and an agent identifier $\mathcal{A}'$. The set of secrets of agent $\mathcal{A}$ is denoted by $\mathcal{S}(\mathcal{K}_\mathcal{A}) = \{(\Phi_1, Bel_1, \mathcal{A}_1), \ldots, (\Phi_n, Bel_n, \mathcal{A}_n)\}$.*

The semantics of a secret is that if agent $\mathcal{D}$ holds the secret $(\Phi, Bel_\mathcal{A}, \mathcal{A}) \in \mathcal{S}(\mathcal{K}_\mathcal{D})$, it does not want that agent $\mathcal{A}$ believes Φ by use of the belief operator $Bel_\mathcal{A}$, i. e. $\Phi \notin Bel_\mathcal{A}(f_\mathcal{A}(\mathcal{K}_\mathcal{D}))$. Since we consider agents that only have local views, we have to base secrecy on the subjective view $f_\mathcal{A}(\mathcal{K}_\mathcal{D})$ agent $\mathcal{D}$ has on the beliefs of another agent $\mathcal{A}$.

Example 5 *We can formalize the secrets of Alfred as $\mathcal{S}(\mathcal{K}_\mathcal{A}) = \{(\mathsf{affair}, Bel_{cred}, \mathcal{B}), (\mathsf{children}, Bel_{skep}, \mathcal{C})\}$ assuming the propositions* affair *and* children *with the obvious meaning.*

The set of secrets is dynamic in the sense that new secrets might be added, weakened, strengthened or removed.

Example 6 *If Alfred gets to know that some agent $\mathcal{D}$ told Carla that he has children he should give up his corresponding secret.*

This leads to a subjective and dynamic notion of secrecy based on the point of view of a single agent which we consider very natural since it reflects how humans treat secrecy every day.

Each agent can perform actions $a \in actions$ in its environment and receives perceptions $p \in percepts$ from the environment. In this work we constrain our considerations to communication acts and, for the sake of clarity and without loss of generality, we focus on just two agents. Since both, actions as well as perceptions imply epistemic changes we generalize actions and perceptions to pieces of information $\tau \in actions \cup percepts$. For the changes to an epistemic state that are implied due to the execution of some action or the incorporation of some perception we define a belief change operator.

Definition 5 (Belief Change Operator) *We assume an operator $\circ$ which changes the epistemic state $\mathcal{K}_\mathcal{A}$ of an agent $\mathcal{A}$ given some information $\tau \in actions \cup percepts$, such that $\mathcal{K}_\mathcal{A} \circ \tau = \mathcal{K}'_\mathcal{A}$.*

The concept of the belief change operator on epistemic states and information is very versatile. A belief change operator has to embrace the type of epistemic state, adequate operators for actions and for perceptions, for different types of sequences of pieces of information, the estimated effects on the epistemic states of other agents and the dynamics of secrets. These tasks call for different types of operations, e. g, update and revision operators (Katsuno and Mendelzon 1994; Lang 2006), but abstract from such subtleties here and assume the proposed operator to deal with these appropriately.

Agent-based epistemic secrecy

In this section we elaborate a subjective, agent based notion of epistemic secrecy based on the setting and framework presented before. We start by formulating our idea of secrecy intuitively and formalize it afterwards. Our intuition of agent based epistemic secrecy is that: *An agent $\mathcal{D}$ preserves secrecy if, from its point of view, none of its secrets Φ that it wants to hide from agent $\mathcal{A}$ is, from $\mathcal{D}$'s perspective, believed by $\mathcal{A}$ after any of $\mathcal{D}$'s actions (given that $\mathcal{A}$ does not believe Φ already).*

This intuitive idea expresses that we want to assure that the secrecy preserving agent always maintains an epistemic state in which it believes that no other agent believes in something that it wants to keep secret. More exactly, it also distinguishes between secrets towards different agents and what it means to it that the information is kept secret. That the agent shall "*always maintain*" as written above means

that for all possible scenarios of communication it acts such that a safe epistemic state is maintained. We formalize the intuitions laid out above in the following definition.

Definition 6 (Safe Epistemic State) *An epistemic state $\mathcal{K}_D$ is* safe *if and only if it holds that $\Phi \notin Bel(f_{\mathcal{A}}(\mathcal{K}_{\mathcal{D}}))$ for all $(\Phi, Bel, \mathcal{A}) \in \mathcal{S}(\mathcal{K}_{\mathcal{D}})$. We denote the set of all safe epistemic states by $\Lambda \subseteq \Omega$.*

An agent reveals information by performing actions, that is, either by communicating with agents directly or indirectly by actions that are observed by other agents. For the sake of simplicity of presentation we restrict action to be communication actions only. An agent preserves secrecy if it does not perform any action that leads to an unsafe epistemic state. The basic model of an agent is that of an entity which in one cycle receives a perception from its environment and performs an action. The perception p received by an agent $\mathcal{X}$ leads to a modification of its epistemic state by its change operator such that $\mathcal{K}'_{\mathcal{X}} = \mathcal{K}_{\mathcal{X}} \circ p$. We model the action $a \in actions$ of an agent to be determined by the current epistemic state $\mathcal{K}_A \in \Omega$ such that an agent is modeled by an action function

$$act : \Omega \to actions.$$

We denote the set of all action functions by $\mathbb{A}$. Actions and perception can be the empty action or empty perception. An agent starts with an initial epistemic state. Based on this and its perceptions it acts in its environment. The new epistemic state is then given by the revision of its current state by the information it gets from its environment in form of a perception and by the information which action it has performed. An agent cycle results in a new epistemic state determined by $\mathcal{K}_{\mathcal{D}} \circ p \circ act_{\mathcal{D}}(\mathcal{K}_{\mathcal{D}} \circ p)$.

Normally, an agent is assumed to get feedback of its actions by observing changes in its environment. This is not applicable here since the effects of actions cause changes in the epistemic states of other agents which cannot be directly observed. In our approach these changes are simulated in the view of one agent on other agents. That is, the change operator updates $\mathcal{D}$'s view on other agents' epistemic states given a performed action by $\mathcal{D}$.

We demand that the actions of an agent do not disclose any secrets. Formally we want that all safe epistemic states $\mathcal{K}_{\mathcal{D}} \in \Lambda$ and for all percepts $p \in percepts$ we have that $\mathcal{K}_{\mathcal{D}} \circ p \circ act_{\mathcal{D}}(\mathcal{K}_{\mathcal{D}} \circ p)$ is safe. We weaken this requirement by assuming a set of initial, safe, epistemic states of an agent and apply the condition above only to all states accessible through the behavior of the agent from an initial state. The set of all possible epistemic states of agent $\mathcal{D}$ is determined by the set of initial epistemic states Λ_0 and all respective successor states for all possible perceptions and corresponding actions of $\mathcal{D}$ i. e.

$$\Omega_{act,percepts}(\Lambda_0) = \{\mathcal{K} \mid \mathcal{K} = \mathcal{K}_0 \circ p_0 \circ act(\mathcal{K}_0 \circ p_0) \circ \ldots, p_i \in percepts, i \in \mathbb{N}_0, \mathcal{K}_0 \in \Lambda_0\}.$$

Based on these considerations we define our requirements to an agent.

Definition 7 (Secrecy Preserving Agent) *Given a set of safe epistemic states Λ_0 and a set of perceptions $percepts$. Let be $\mathcal{D} \in \mathfrak{A}$ be an agent, $act_{\mathcal{D}}$ its action function. We call $\mathcal{D}$* secrecy *preserving if and only if for all $\mathcal{K}_{\mathcal{X}} \in \Omega_{act,percepts}(\Lambda_0)$ it holds that $\mathcal{K}_{\mathcal{X}}$ is safe.*

Therefore we require the agent to act such that secrecy is not violated, if it is not violated yet. The latter is expressed by requiring $\mathcal{K}_{\mathcal{A}}$ to be an element of the set of safe epistemic states Λ. That is, we formulate secrecy as a requirement on the function act. This means that the agent has to take care that it does not perform any action which violates secrecy. In the next section we show how our framework and notion of secrecy can be instantiated for the runs-and-system framework and how it relates to other notions of secrecy.

Instantiation for the Runs-and-Systems Framework

The body of work on secrecy is large and diverse. The most general and accepted framework for definitions of secrecy in multiagent systems is the one of Halpern and O'Neill (Halpern and O'Neill 2008). It generalizes several other notions of secrecy and forms a basis for discussions of general issues of secrecy. In (Biskup and Tadros 2010) Biskup and Tadros generalized the notions of Halpern and O'Neil through their policy-based notion of secrecy. In this section we present the runs-and-systems framework and some notions of secrecy based on it. Afterwards we formulate an instantiation of our framework that captures the ideas of the runs-and-systems framework and show how we can express other notions of secrecy within this framework. Based on the formalization of other notions of secrecy within our framework we can show presumptions and properties of those and further possibilities opened by our framework.

Secrecy in runs-and-sytems

Halpern and O'Neill base their work on the runs-and-systems framework (Fagin et al. 1995). A *system* $\mathcal{R}$ is a set of runs r whereby a run is a sequence of global states. Global states of a system are identified using the two dimensions of runs r and time-points m such that a point is given by (r, m) and the corresponding global state is denoted by $r(m)$. A global state consists of all local states of all agents $\mathfrak{A} = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$, such that $r(m) = (s_{\mathcal{X}_1}, s_{\mathcal{X}_2}, \ldots, s_{\mathcal{X}_n})$. The local state of agent $\mathcal{X}$ in a global state $r(m)$ is given by $r_{\mathcal{X}}(m)$. The set of all points of a system $\mathcal{R}$ is denoted by $\mathcal{PT}(\mathcal{R})$. The view of the system, also called $\mathcal{X}$*-information set*, of an agent $\mathcal{X} \in \mathfrak{A}$ at point (r, m) on system $\mathcal{R}$ is defined as the set of points that it considers possible in (r, m): $\mathcal{K}_{\mathcal{X}}(r, m) = \{(r', m') \in \mathcal{PT}(\mathcal{R}) \mid r'_{\mathcal{X}}(m') = r_{\mathcal{X}}(m)\}$. Here, we also use an equivalent definition with a local state $s_{\mathcal{X}}$ as argument instead of a global state (r, m), i. e. $\mathcal{K}_{\mathcal{X}}(s_{\mathcal{X}}) = \{(r', m') \in \mathcal{PT}(\mathcal{R}) \mid r'_{\mathcal{X}}(m') = s_{\mathcal{X}}\}$. This formulation seems more intuitive from the agent's the perspective.

In (Halpern and O'Neill 2008) several qualitative notions of secrecy are presented which all base on the principle that an agent $\mathcal{D}$[2] maintains secrecy with respect to an agent $\mathcal{A}$ if

[2] In (Halpern and O'Neill 2008) the attacking agent $\mathcal{A}$ is denoted by i and the defending $\mathcal{D}$ by j.

and only if for all possible points $(r, m) \in \mathcal{PT}(\mathcal{R})$ agent $\mathcal{A}$ cannot rule out secrecy relevant information $I_\mathcal{D}$ of $\mathcal{D}$. The idea is that $I_\mathcal{D}$ describes a (secrecy-)relevant property of agent $\mathcal{D}$ that needs to be protected. The strongest notion is the one of total secrecy in which $I_\mathcal{D}$ is the set of all possible local states of $\mathcal{D}$.

Definition 8 (Total Secrecy) *Agent $\mathcal{D}$ maintains* total secrecy *with respect to $\mathcal{A}$ in system $\mathcal{R}$ if, for all points (r, m) and $(r', m') \in \mathcal{PT}(\mathcal{R})$ $\mathcal{K}_\mathcal{A}(r, m) \cap \mathcal{K}_\mathcal{D}(r', m') \neq \emptyset$.*

This notion of secrecy inhibits any information flow and, as stated in (Halpern and O'Neill 2008), "for almost any imaginable system, it is, in fact, too strong to be useful."

Example 7 *In our example, total secrecy would imply that Beatriz is not allowed to belief that Alfred knows his own name since all information is declared secrecy relevant.*

Several weaker notions of secrecy are presented in (Halpern and O'Neill 2008) that restrict the amount of relevant information $I_\mathcal{D}$ of agent $\mathcal{D}$. In *total f-secrecy* relevant information is defined by relevant values of $\mathcal{D}$ such that not the entire local state is relevant. In *C-secrecy* a $\mathcal{A}$-allowability function explicitly defines a set of points $\mathcal{A}$ is allowed to rule out in each point.

The most general definition of secrecy is *policy-based secrecy* presented in (Biskup and Tadros 2010). In policy-based secrecy the relevant information of agent $\mathcal{D}$ is defined by a set $\mathcal{I}_\mathcal{D}$ of sets of $\mathcal{D}$-information sets $I_\mathcal{D}$, i. e. $\mathcal{I}_\mathcal{D} = \{I^1_\mathcal{D}, \ldots, I^l_\mathcal{D}\}$. The construction via sets of information sets allows for a more expressive formulation of relevant information as it can be expressed that out of a set $I_\mathcal{D}$ some local states can be excluded but not all which corresponds to the disjunction of information. That is, every set of $\mathcal{D}$-information sets characterizes some relevant property and the set of those all relevant properties of agent $\mathcal{D}$. This is formalized by a $\mathcal{D}$-possibility policy.

Definition 9 ($\mathcal{D}$-Possibility Policy) *A $\mathcal{D}$-*possibility policy *policy is a function: $\mathcal{PT}(\mathcal{R}) \to \mathcal{P}(\mathcal{P}(\mathcal{P}(\mathcal{PT}(\mathcal{R}))))$ such that $policy(r, m) := \{\mathcal{I}^1_\mathcal{D}, \mathcal{I}^2_\mathcal{D}, \ldots\}$ contains sets $\mathcal{I}_\mathcal{D}$ of $\mathcal{D}$-information sets.*

Policy-based secrecy is now defined as follows.

Definition 10 (Policy-based Secrecy) *If policy is a $\mathcal{D}$-possibility policy, agent $\mathcal{D}$ maintains policy-based secrecy with respect to agent $\mathcal{A}$ in $\mathcal{R}$ if, for all points $(r, m) \in \mathcal{PT}(\mathcal{R})$ and for all $\mathcal{I}^k_\mathcal{D} \in policy(r, m)$:*

$$\mathcal{K}_\mathcal{A}(r, m) \cap \bigcup_{I_\mathcal{D} \in \mathcal{I}^k_\mathcal{D}} I_\mathcal{D} \neq \emptyset$$

That is, no property of $\mathcal{D}$ characterized by an $\mathcal{I}^k_\mathcal{D}$ should be ruled out by agent $\mathcal{A}$. Various other notions of secrecy like total secrecy, C-secrecy and total f-secrecy are shown to be special cases of policy-based secrecy in (Biskup and Tadros 2010).

Epistemic secrecy

In the following we elaborate the epistemic view on the runs-and-systems approach and how the epistemic state of an agent within the system could be in order to reflect agent-based epistemic secrecy. In particular, each agent has to be aware of the information it wants to keep secret and about the beliefs and reasoning of other agents. Afterwards we relate our notion of epistemic secrecy to policy-based secrecy.

We adapt the compound version of an epistemic state of Definition 2 to define the epistemic state of an agent in a system. In the following we define each component with respect to the runs-and-systems framework.

The belief base B of agent $\mathcal{X}$ contains a base representation of the beliefs it currently holds. Information in the runs-and-systems framework is represented by sets of states, called *information-sets* $I \subseteq \mathcal{PT}(\mathcal{R})$. Thus, the language under consideration is given by the power-set of the set of points of the given system, i. e. $\mathcal{L}_{BS} = \mathcal{P}(\mathcal{PT}(\mathcal{R}))$. The belief base is a set of points $B \subseteq \mathcal{PT}(\mathcal{R}) = \mathcal{L}_V$.

The information about a particular agent $\mathcal{X}$ is correspondingly expressed by a set of local states $r_\mathcal{X}(m)$ of $\mathcal{X}$. Each information-set $I \subseteq \mathcal{PT}(\mathcal{R})$ contains information about agent $\mathcal{X}$, the local information-set $I_\mathcal{X}$, given by

$$I_\mathcal{X}(I) = \{r_\mathcal{X}(m) \mid (r, m) \in I\}.$$

It holds for any point (r, m) that $I_\mathcal{X}(\mathcal{K}_\mathcal{X}(r, m)) = \{r_\mathcal{X}(m)\}$. This captures the special case of the local state of agent $\mathcal{X}$ being uniquely determined which corresponds to certainty of the information. To allow for uncertainty, we generalize the $\mathcal{K}$ operator to sets of local states $I_\mathcal{X}$ by $\mathcal{K}_\mathcal{X}(I_\mathcal{X}) = \bigcup_{s \in I_\mathcal{X}} \mathcal{K}_\mathcal{X}(s)$. Obviously for any point (r, m) it holds that $I_\mathcal{X}(\mathcal{K}_\mathcal{X}(I_\mathcal{X})) = I_\mathcal{X}$. The set of views $\mathcal{V}$ of agent $\mathcal{D}$ on the other agents is generated by sets of local states of the respective agent, i. e. $\mathcal{V} = \{\mathcal{V}_{\mathcal{X}_1}, \ldots, \mathcal{V}_{\mathcal{X}_n}\}$ with $\mathcal{V}_{\mathcal{X}_i} = \mathcal{K}_{\mathcal{X}_i}(\{s^1_{\mathcal{X}_i}, \ldots, s^l_{\mathcal{X}_i}\}) \subseteq \mathcal{PT}(\mathcal{R})$. A special case is the one of $I_\mathcal{A}(\mathcal{V}_\mathcal{A})$ being a singleton, that is, agent $\mathcal{D}$ has no uncertainty about the state of agent $\mathcal{A}$. The natural definition of the view of agent $\mathcal{D}$ in state $s_\mathcal{D}$ on $\mathcal{A}$ would be $\mathcal{V}_\mathcal{A} = I_\mathcal{A}(\mathcal{K}_\mathcal{D}(s_\mathcal{D}))$.

Secrets are triples of the form $(I, Bel_\mathcal{R}, \mathcal{X})$. The information to be kept secret I is, as all information in the runs-and-systems framework, represented as a set of states, i. e. $I \subseteq \mathcal{PT}(\mathcal{R})$. Clearly, we have $\mathcal{X} \in \mathfrak{A}$. For the adequate definition of a belief operator we have to clarify the concept of inference in the runs-and-systems framework. The inference notion used for secrecy based on the runs-and-systems approach states that an agent $\mathcal{X}$ can infer all information for which it can exclude that it does not hold. Formally that is, all information represented by an information set $I \subseteq \mathcal{PT}(\mathcal{R})$ can be inferred for which $(\mathcal{PT}(\mathcal{R}) \setminus I) \cap \mathcal{K}(s_\mathcal{X}) = \emptyset$. This brings us to the following definition of a belief operator.

Definition 11 *Let $\mathcal{R}$ be a system and $s_\mathcal{X}$ a local state of agent $\mathcal{X} \in \mathfrak{A}$. We define the $\mathcal{R}$* belief operator *as*

$$Bel_\mathcal{R}(s_\mathcal{X}) = \{I \subseteq \mathcal{PT}(\mathcal{R}) \mid (\mathcal{PT}(\mathcal{R}) \setminus I) \cap \mathcal{K}_\mathcal{X}(s_\mathcal{X}) = \emptyset\}.$$

That is the belief operator of agents is uniquely defined and does not allow for alternative definitions as skeptical or credulous inference. Given a specific local state, there is no notion of uncertainty for the respective agent. From the perspective of agent $\mathcal{D}$ this means that it knows the belief operator of agent $\mathcal{A}$ and that it does not want it to be able to infer

any secret using this operator. Uncertainty, however, arises from $\mathcal{D}$ not knowing the exact local state of agent $\mathcal{A}$. Then $\mathcal{D}$ considers $S_\mathcal{A} = \{s^1_\mathcal{A}, \ldots, s^l_\mathcal{A}\}$ a set of local states of agent $\mathcal{A}$ to be the actual local state of $\mathcal{A}$.

Definition 12 *Let $\mathcal{R}$ be a system and $S_\mathcal{X} = \{s^1_\mathcal{X}, \ldots, s^l_\mathcal{X}\}$ a set of local states of agent $\mathcal{X} \in \mathfrak{A}$. We define the* sceptical belief operator *as*

$$\begin{aligned} Bel_{\mathcal{R},s}(S_\mathcal{X}) &= \{I \subseteq \mathcal{PT}(\mathcal{R}) \mid \forall s_\mathcal{X} \in S_\mathcal{X}, \\ &\quad (\mathcal{PT}(\mathcal{R}) \setminus I) \cap \mathcal{K}_\mathcal{X}(s_\mathcal{X}) = \emptyset\}. \end{aligned}$$

Example 8 *Let I_{affair} denote the set of states in which Alfred has an affair then he wants that Beatriz, who he believes to be in state $s_\mathcal{B}$, is not able to exclude all states in which he does not have an affair. This is the case exactly if $I_{\mathsf{affair}} \notin Bel_{\mathcal{R},s}(S_\mathcal{A}, Bel_\mathcal{R}(\{s_\mathcal{B}\}))$.*

The scenario just described changes if an agent $\mathcal{D}$ has uncertain information about the current state of a possible attacker $\mathcal{A}$. In the runs-and-systems framework this is the case if $f_\mathcal{A}(\mathcal{K}_\mathcal{D})$ has more than one element. Besides the skeptical operator defined above we can define a credulous operator by replacing the all quantification by an existential one.

Definition 13 *Let $\mathcal{R}$ be a system and $S_\mathcal{X} = \{s^1_\mathcal{X}, \ldots, s^l_\mathcal{X}\}$ a set of local states of agent $\mathcal{X} \in \mathfrak{A}$. We define the* credulous belief operator *as*

$$\begin{aligned} Bel_{\mathcal{R},c}(S_\mathcal{X}) &= \{I \subseteq \mathcal{PT}(\mathcal{R}) \mid \exists s_\mathcal{X} \in S_\mathcal{X}, \\ &\quad (\mathcal{PT}(\mathcal{R}) \setminus I) \cap \mathcal{K}_\mathcal{X}(s_\mathcal{X}) = \emptyset\}. \end{aligned}$$

Example 9 *Let I_{affair} denote the set of states in which Alfred has an affair then he wants that Beatriz. He is not sure whether Beatriz heard him talking on the phone last night, $s^p_\mathcal{B}$, or not, $s^{\neg p}_\mathcal{B}$. As Alfred is cautious, he wants that in both cases Beatriz does not believe that he has an affair and demands $I_{\mathsf{affair}} \notin Bel_{\mathcal{R},c}(S_\mathcal{A}, Bel_\mathcal{R}(\{s^{\neg p}_\mathcal{B}, s^p_\mathcal{B}\}))$.*

On the other hand, if he is not sure if Carla heard him talking to his son on the phone, represented by Carla's possible states $s^p_\mathcal{C}$ and $s^{\neg p}_\mathcal{C}$, he might not care too much. In this case he would only require that $I_{\mathsf{children}} \notin Bel_{\mathcal{R},s}(\{s^{\neg p}_\mathcal{C}, s^p_\mathcal{C}\})$.

Based on this epistemic view on and formalization of the runs-and-systems approach we can define an *epistemic system* $\mathcal{R}_\mathcal{E}$ for a given system R such that each agent has all information it needs to be aware of and reason about secrecy.

Definition 14 (Epistemic System) *Let $\mathcal{R}$ be a system for a set of agents $\mathfrak{A} = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$. The* epistemic system $\mathcal{R}_\mathcal{K}(\mathcal{R})$ *for $\mathcal{R}$ is defined as*

$$\begin{aligned} \mathcal{R}_\mathcal{K}(\mathcal{R}) = \{r_\mathcal{K} \quad \mid \quad & r_{\mathcal{K},\mathcal{X}}(m) = \mathcal{K}_\mathcal{X}(r,m), \\ & \mathcal{X} \in \mathfrak{A}, (r,m) \in \mathcal{PT}(\mathcal{R})\} \end{aligned}$$

with $\mathcal{K}_\mathcal{X}(r,m) =$

$$\begin{aligned} &\langle r_\mathcal{X}(m), \{(I_{\mathcal{X}_1,1}, Bel_\mathcal{R}, \mathcal{X}_1), \ldots, (I_{\mathcal{X}_1,k}, Bel_\mathcal{R}, \mathcal{X}_1), \\ &\qquad \ldots, (I_{\mathcal{X}_n,k}, Bel_\mathcal{R}, \mathcal{X}_n)\}, \{\mathcal{V}_{\mathcal{X}_1}, \ldots, \mathcal{V}_{\mathcal{X}_n}\}\rangle \end{aligned}$$

The information that has to be specified for the epistemic system is, for each agent, the set of secrets and the set of views for all points in the system $\mathcal{R}$. Current definitions of secrecy are less expressive in the sense of definition of secrets. The most expressive definition of secrecy is the one of policy-based secrecy. In the following we show how we can express policy-based secrecy in an epistemic agent-based system.

Properties of the Instantiation

There are several differences in the concept of secrecy and ours. Most striking is the difference in the definition of secret information. The information to be kept secret in secrecy is implicit. The relevant information is specified and does not represent secrets but information not to be ruled out, that is, information to be considered possible by agent $\mathcal{A}$. In comparison to the notion of secrets in our framework this means that all information is to be kept secret in which some relevant information is ruled out. Formally, this means that if $\mathcal{I} = \{I_1, \ldots I_l\} \subseteq \mathcal{P}(\mathcal{PT}(\mathcal{R}))$ characterizes relevant information, the set of secret information is given by

$$\Gamma(\mathcal{I}) = \{I \mid \exists I' \in \mathcal{I}, I = \mathcal{PT}(\mathcal{R}) \setminus I'\}.$$

This set of secrets in form of information-sets forms the basis for the secrets of agent $\mathcal{D}$. Secrecy is not only defined for a single agent but also with respect to one single other agent while our notions consider all other agents and the respective secrets at the same time. Therefore we restrict our framework to sets of secrets which are with respect to the same agent, i. e. $\mathcal{A}$. The information to be kept secret in policy-based secrecy, as in other the notions of secrecy is local information about agent $\mathcal{D}$. This is expressed by the condition of all sets $\mathcal{I}_k$ being sets of $\mathcal{D}$-information sets. For the case of total secrecy the relevant information is given by the set of all possible local states of $\mathcal{D}$, i. e.

$$\mathcal{I}_{total} = \{\mathcal{K}_\mathcal{D}(r,m) \mid (r,m) \in \mathcal{PT}(\mathcal{R}))\}.$$

Given some $\mathcal{D}$-possibility policy $p(r,m)$ for a point $(r,m) \in \mathcal{PT}(\mathcal{R})$ we get the relevant information by the union over each set of information-sets:

$$\mathcal{I}_{p(r,m)} = \{ \bigcup_{I \in \mathcal{I}_k} I \mid \mathcal{I}_k \in p(r,m)\}.$$

The following definition relates agent-based epistemic secrecy to policy-based secrecy by transforming a given system, pair of agents and policy into a system in which agents have epistemic states as defined in Definition 2.

Definition 15 *For a given system $\mathcal{R}$, agents $\mathcal{D}$ and $\mathcal{A}$, and policy $policy(r,m)$ for all $(r,m) \in \mathcal{PT}(\mathcal{R}))$ we define the system $\mathcal{R}_\mathcal{K}(\mathcal{R})$ as*

$$\mathcal{R}_\mathcal{K}(\mathcal{R}) = \{r_\mathcal{K} \mid r_{\mathcal{K},\mathcal{X}}(m) = \mathcal{K}_\mathcal{X}(r,m), \mathcal{X} \in \mathfrak{A}\}$$

with

$$\mathcal{K}_\mathcal{X}(r,m) = \begin{cases} \langle r_\mathcal{X}(m), \{(I, Bel_{\mathcal{R},c}, \mathcal{A}) \mid \\ I \in \Gamma(\mathcal{I}_{p(r,m)})\}, \{\{r_\mathcal{A}(m)\}\}\rangle & \text{if } \mathcal{X} = \mathcal{D} \\ \langle r_\mathcal{X}(m), \emptyset, \emptyset\rangle & \text{else} \end{cases}$$

The set of secrets is given by the secrecy-relevant elements of $policy(r,m)$, all secrets are protected with respect to the possibilistic belief operator $Bel_{\mathcal{R},c}$. Since all secrecy definitions based on the runs-and-systems approach define secrecy only with respect to one single agent, i. e. all secrets in $\mathcal{S}$ are for agent $\mathcal{A}$. Accordingly only the view of agent $\mathcal{D}$ on agent $\mathcal{A}$ is relevant and the set of views consists of only the view towards agent $\mathcal{A}$. This set is a singleton such that agent $\mathcal{D}$ has a complete view of agent $\mathcal{A}$. With this instantiation we can show the following result.

Proposition 1 *Given a system $\mathcal{R}$, agents $\mathcal{D}$ and $\mathcal{A}$, and policy $policy(r, m)$ for all $(r, m) \in \mathcal{PT}(\mathcal{R})$. Agent $\mathcal{D}$ maintains policy-based secrecy with respect to agent $\mathcal{A}$ in $\mathcal{R}$ if and only if, for all points $(r_\mathcal{K}, m) \in \mathcal{PT}(\mathcal{R}_\mathcal{K})$ $r_{\mathcal{K},\mathcal{D}}(m)$ is safe.*

Proof 1 *Assume that $\mathcal{D}$ maintains policy-based secrecy with respect to $\mathcal{A}$. We fix some arbitrary point $(r, m) \in \mathcal{PT}(\mathcal{R})$ and $\mathcal{I} \in \mathcal{I}_{p(r,m)}$. It is $\mathcal{I}_{p(r,m)} = \{\bigcup_{I \in \mathcal{I}_k} I \mid \mathcal{I}_k \in p(r, m)\}$. By definition of policy-based secrecy it holds that $\mathcal{K}_\mathcal{A}(r, m) \cap \mathcal{I} \neq \emptyset$. We have $\mathcal{PT}(\mathcal{R}) \setminus \mathcal{I} \in \Gamma(\mathcal{I}_{p(r,m)})$. Now suppose that $(\mathcal{PT}(\mathcal{R}) \setminus \mathcal{I}) \in Bel_{\mathcal{R},s}(\{r_\mathcal{A}(m)\})$. Then it follows that $\mathcal{PT}(\mathcal{R}) \setminus (\mathcal{PT}(\mathcal{R}) \setminus \mathcal{I}) \cap \mathcal{K}_\mathcal{A}(r, m) = \emptyset$ that is $\mathcal{I} \cap \mathcal{K}_\mathcal{A}(r, m) = \emptyset$, in contradiction to the assumption.*

Assume $r_{\mathcal{K},\mathcal{D}}(m)$ is safe for all points $(r_\mathcal{K}, m) \in \mathcal{PT}(\mathcal{R}_\mathcal{K})$. We fix some arbitrary point $(r, m) \in \mathcal{PT}(\mathcal{R})$ again. Then for each $\mathcal{I}' \in \Gamma(\mathcal{I}_{p(r,m)})$ it holds that $\mathcal{I}' \notin Bel_{\mathcal{R},s}(\{r_\mathcal{A}(m)\})$. From this follows that $\mathcal{PT}(\mathcal{R}) \setminus \mathcal{I}' \cap \mathcal{K}_\mathcal{A}(r, m) \neq \emptyset$. Since for each $\mathcal{I} \in \mathcal{I}_{p(r,m)}$ we have $\mathcal{PT}(\mathcal{R}) \setminus \mathcal{I} \in \Gamma(\mathcal{I}_{p(r,m)})$ it also holds that $\mathcal{PT}(\mathcal{R}) \setminus (\mathcal{PT}(\mathcal{R}) \setminus \mathcal{I}) \cap \mathcal{K}_\mathcal{A}(r, m) \neq \emptyset$ and therefore $\mathcal{I} \cap \mathcal{K}_\mathcal{A}(r, m) \neq \emptyset$ as desired.

We just showed how we can transform any given system and policy into a system in which the agent under consideration has knowledge about the policy and a view on the considered attacking agent. This result shows that by the definition of a policy for one single agent and with respect to one other agent little information is given about the whole system. This gets evident by the fact that all other agent's secrets and views are not specified, as only one agent $\mathcal{D}$ is modeled and that only partially. It also gets evident, that the defending agent $\mathcal{D}$ needs to have complete information about the attacking agent $\mathcal{A}$ in order to satisfy secrecy. This is already unrealistic in this asymmetric setting in which $\mathcal{D}$ is the only agent which has secrets. In many realistic settings of multiagent systems, agent $\mathcal{A}$ is designed symmetrically and should be able to have secrets with respect to $\mathcal{D}$. Both agents should be able to preserve secrecy at the same time. This is impossible if both have a complete view on the other.

The formulation of secrets in policy-based secrecy is unintuitive as we need to specify the information an attacking agent shall believe. Hence if we consider our running example we have.

Example 10 *Beatriz should not be able to rule out that Alfred does not have an affair. Hence $\neg$`affair` is security relevant information.*

The notion of security-relevant is somewhat the complement of our notion of secrets. If some proposition a shall be kept secret $\neg a$ is security-relevant information. We can model this interpretation in our framework as well by using the belief operator

$$Bel_\mathcal{R}(s) = \{X \mid X \subseteq \mathcal{PT}(\mathcal{R}) \setminus \mathcal{K}(s)\}.$$

This operator makes negative inferences which results in the set of information sets representing information the agent does not consider possible. Furthermore, secrecy restricts secrets to be $\mathcal{D}$*-local* which means that it is a property of $\mathcal{D}$. This means that an agent can only keep secret that it beliefs in something. Consider the following example.

Example 11 *Dave is married and does not have an affair. Still he does not want his jealous wife to believe that he has an affair.*

This counterfactual secret cannot be represented on secrecy but in our framework since we do not restrict the secret information to be $\mathcal{D}$-local.

Agent-based Runs

Up to now we assumed a system $\mathcal{R}$ to be given. The agent based notion of secrecy presented above defines secrecy based on the behavior of an agent stating when an agent defined by some agent function preserves secrecy. Hence agent based secrecy is a property of an agent function. That is what we are interested in, agent functions that preserve secrecy. In the following we use the action function defined previously which describes the behavior of agents to construct a system.

A system is a set of runs and each run describes the evolution of the system in time. Hereby, also the evolution of each local state is described. Formally a run is defined as a function $r : \mathbb{N}_0 \to states$ from discrete time points to the set of global states $states$. In our framework the set of local states of the agents $\mathfrak{A} = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$ is given by the set of epistemic states Ω and the set of global states consequently by

$$\mathbb{E} = \{(\mathcal{K}_{\mathcal{X}_1}, \ldots, \mathcal{K}_{\mathcal{X}_n}) \mid \mathcal{K}_{\mathcal{X}_i} \in \Omega, \mathcal{X}_i \in \mathfrak{A}, 1 \leq i \leq n\}.$$

For each local state $\mathcal{K}_\mathcal{X}$ the action function $act_\mathcal{X}(\mathcal{K}_\mathcal{X})$ determines the next action of agent $\mathcal{X}$. An assignment of an action function to each agent is denoted by $Act_\mathfrak{A} = (act_{\mathcal{X}_1}, \ldots, act_{\mathcal{X}_n})$. Each agent performs an action in each cycle and the aggregated actions $\alpha = (a_1, \ldots, a_n)$ form the perception $p_{\alpha,\mathcal{X}}$ of each agent in the next cycle. We leave out details about communication and visibility of actions here. The change operator $\circ$ incorporates the new perceptions and executed actions into the current epistemic state and hereby determines the successor state. It should be stressed that besides the $\circ$ operator not only changes the belief base of the agent but also adapts the view on other agents and the sets of secrets. The adaption of the view of other agents reflects the changes $\mathcal{D}$ supposes that its actions have on the epistemic state of other actions and the information $\mathcal{D}$ gets about other agents by observing their actions, e. g. by applying techniques from (Nittka and Booth 2008). Details about the change operator are out of the focus of this paper.

Given the explanations above we can completely and uniquely describe the execution of a system and define an agent based epistemic system.

Definition 16 (Agent-based epistemic system) *Let $\mathbb{E} = \{\mathcal{K}_{\mathcal{X}_1}, \ldots, \mathcal{K}_{\mathcal{X}_n}\}$ be a set of safe epistemic states with $\mathcal{K}_{\mathcal{X}_i} \in \Lambda_0$, $1 \leq i \leq n$ and $Act_\mathfrak{A}$ an action function assignment for all agents.*

- *The run $r_{\mathbb{E},Act_\mathfrak{A}}$ starts with the initial state $r_{\mathbb{E},Act_\mathfrak{A}}(0) = \{\mathcal{K}_{\mathcal{X}_1}, \ldots, \mathcal{K}_{\mathcal{X}_n}\}$.*
- *$r_{\mathbb{E},Act_\mathfrak{A}}(m) = \{r_{\mathcal{X}_1}(m-1) \circ p_{\mathcal{X}_1}(m-1) \circ a_{\mathcal{X}_1}(m-1), \ldots, r_{\mathcal{X}_n} \circ p_{\mathcal{X}_n}(m-1) \circ a_{\mathcal{X}_n}(m-1)\}$*
  - *with $a_{\mathcal{X}_i}(m) = act_{\mathcal{X}_i}(r_{\mathcal{X}_1}(m) \circ p_1(m))$*

The agent based epistemic system $\mathcal{R}_{\mathcal{E}}$ *is the set of runs* $r_{\mathbb{E},Act_{\mathfrak{A}}}$ *for all possible initial epistemic states* $\mathbb{E}$ *and assignments* $Act_{\mathfrak{A}}$*. We denote the set of runs in which agent* $\mathcal{X}$ *has behavior* $act_{\mathcal{X}}$ *by*

$$\mathcal{R}_{act_{\mathcal{X}}} := \{r_{\mathbb{E},Act_{\mathfrak{A}}} \in \mathcal{R} \mid Act_{\mathfrak{A}}(\mathcal{X}) = act_{\mathcal{X}})\}.$$

The system $\mathcal{R}_{act_{\mathcal{X}}}$ is relevant for our means since it represents all runs in which $\mathcal{X}$ has a certain action function and the other agent have all possible action functions for all possible initial epistemic states. We can show that the construction of epistemic systems characterizes our notion of agent-based epistemic secrecy.

Proposition 2 *Given a set of perceptions* $percepts$*, a set of initial epistemic states* Λ*. An agent* $\mathcal{X}$ *characterized by behavior* $act_{\mathcal{X}}$ *is secrecy preserving if and only if* $r_{\mathcal{X}}(m)$ *is safe for all points* $(r, m) \in \mathcal{PT}(\mathcal{R}_{\mathcal{E},act_{\mathcal{X}}})$*.*

Proof 2 *We have to show that* $\{r_{\mathcal{X}}(m) \mid (r, m) \in \mathcal{PT}(\mathcal{R}_{\mathcal{E},act_{\mathcal{X}}})\} = \Omega_{act,percepts}(\Lambda_0)$*. If* $(r, m) \in \mathcal{PT}(\mathcal{R}_{\mathcal{E},act_{\mathcal{X}}})$ *then* $r_{\mathcal{X}}(0) \in \Lambda_0$ *and agent* $\mathcal{X}$ *has behavior* $act_{\mathcal{X}}$*. For each* m *it then holds that* $r_{\mathcal{X}}(m) = r_{\mathcal{X}}(0) \circ p_0 \circ act(r_{\mathcal{X}}(0) \circ p_0) \circ \cdots \circ p_{m-1} \circ act(r_{\mathcal{X}}(0) \circ \cdots \circ p_{m-1})$*. Since* $p_i \in percepts$ *it holds that* $r_{\mathcal{X}}(m) \in \Omega_{act,percepts}(\Lambda_0)$*.*

For each $\mathcal{K}_{\mathcal{X}} \in \Omega_{act,percepts}(\Lambda_0)$ *it holds that it is the result of the modifications to it resulting from a sequence of received perceptions and performed actions:* $\mathcal{K}_{\mathcal{X}} = \mathcal{K}_{\mathcal{X}(0)} \circ p_0 \circ act(r_{\mathcal{X}}(0) \circ p_0) \circ \cdots \circ p_{m-1} \circ act(\mathcal{K}_{\mathcal{X}(0)} \circ \cdots \circ p_{m-1})$ *for some sequence* $(p_0, \ldots, p_{m-1})$ *with* $p_i \in percepts$*,* $1 \leq i < m$ *and* $\mathcal{K}_{\mathcal{X}(0)} \in \Lambda_0$*. Therefore* $\mathcal{K}_{\mathcal{X}} = r_{\mathcal{X}}(m)$ *for some* $(r, m) \in \mathcal{PT}(\mathcal{R}_{\mathcal{E},act_{\mathcal{X}}})$*.*

Hence we have shown how an agent-based epistemic system can be constructed given a change operator and the specification of action functions. This also formalized the relation of secrecy preserving action functions to secrecy in the resulting system. The other way around, for each system $\mathcal{R}_{\mathcal{K}}(\mathcal{R})$ and change operator $\circ$ there exists a set of behaviors Act such that $\mathcal{R}_{\mathcal{K},Act,\circ} = \mathcal{R}_{\mathcal{K}}(\mathcal{R})$. This characterizes a set of action functions for agent $\mathcal{D}$ and in particular action functions for which it preserves secrecy.

Discussion

To our knowledge no similar approach to secrecy from a subjective, epistemic perspective of an agent has been put forward so far. Related notions of secrecy in multiagent system are formulated from the global perspective of the entire system. We have already extensively discussed and shown the relation to the work on secrecy in the runs-and-systems framework, especially (Halpern and O'Neill 2008) and (Biskup and Tadros 2010). We consider these works as a good basis for comparisons of notions of secrecy for which other relations have been shown already. In particular it has been shown in (Halpern and O'Neill 2008) that separability (McLean 1994) and generalized non-interference (McCullough 1987) are stricter than C-secrecy and total f-secrecy which are special cases of policy-based secrecy. In (Biskup and Tadros 2010) the close relation of policy-based secrecy to opaqueness properties of function views as defined in (Hughes and Shmatikov 2004) is shown.

The aim of this work is to define a framework for the specification of secrecy preserving agents. In particular this shall be used for the construction of agent models and for the application in agent systems that preserve (epistemic) secrecy. On this side, related work is the one of Biskup et al. on controlled query evaluation (CQE) (Biskup 2010; Biskup and Weibert 2007). In CQE database queries are controlled via a censor function. The notion of secrecy underlying CQE is formalized by policy-based secrecy as shown in (Biskup and Tadros 2010). If we model the database query scenario as a simple multiagent system, the censor function corresponds to an action function that leads to a secrecy preserving agent. In (Biskup, Kern-Isberner, and Thimm 2008) it has been discussed how CQE techniques can be used for preserving secrecy in multiagent negotiation.

In this work we approached the topic of secrecy from the perspective of an epistemic agent. From this perspective information is uncertain and incomplete, secrets are specific to other agent and vary in strength. The agent is interested in achieving its goals and while doing this has to take care not to disclose information it does not want to be disclosed. We presented a general epistemic agent model in which we define secrets and what it means to an agent to preserve secrecy. To this end we introduced action functions characterizing the behavior of an agent and a change operator which adapts the beliefs of the agent, its secrets and views on other agents. Starting from this framework we discussed the properties of notions of secrecy based on the runs-and-systems approach. We formally showed how policy-based secrecy relates to our approach. These results imply further relations to various other notions of secrecy. It turned out that we have to restrict our framework in various ways to make it compatible with the ones based on runs-and-systems.

We see our framework as a good starting point for the development of secrecy preserving agents and implementations of those. This drives our current and future work.

Acknowledgements: This work has been supported by the DFG, Collaborative Research Center SFB876, project A5. (`http://sfb876.tu-dortmund.de`)

References

[Biskup and Tadros 2010] Biskup, J., and Tadros, C. 2010. Policy-based secrecy in the runs & systems framework and controlled query evaluation. In *Proceedings of the 5th International Workshop on Security (IWSEC 2010)*, 60–77. Information Processing Society of Japan (IPSJ).

[Biskup and Weibert 2007] Biskup, J., and Weibert, T. 2007. Keeping secrets in incomplete databases. *International Journal of Information Security* online first.

[Biskup, Kern-Isberner, and Thimm 2008] Biskup, J.; Kern-Isberner, G.; and Thimm, M. 2008. Towards enforcement of confidentiality in agent interactions. In Pagnucco, M., and Thielscher, M., eds., *Proceedings of the 12th International Workshop on Non-Monotonic Reasoning (NMR'08)*, 104–112. Sydney, Australia: University of New South Wales, Technical Report No. UNSW-CSE-TR-0819.

[Biskup 2010] Biskup, J. 2010. Usability confinement of server reactions: Maintaining inference-proof client views

by controlled interaction execution. In Kikuchi, S.; Sachdeva, S.; and Bhalla, S., eds., *Databases in Networked Information Systems*, volume 5999 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. 80–106.

[Delgrande et al. 2008] Delgrande, J.; Schaub, T.; Tompits, H.; and Woltran, S. 2008. Belief revision of logic programs under answer set semantics. In Brewka, G., and Lang, J., eds., *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*, 411–421. AAAI Press.

[Fagin et al. 1995] Fagin, R.; Halpern, J. Y.; Moses, Y.; and Vardi, M. Y. 1995. *Reasoning about Knowledge*. MIT Press.

[Gärdenfors and Makinson 1988] Gärdenfors, P., and Makinson, D. 1988. Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*, 83–95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

[Halpern and O'Neill 2008] Halpern, J. Y., and O'Neill, K. R. 2008. Secrecy in multiagent systems. *ACM Transactions on Information and System Security (TISSEC)* 12:5:1–5:47.

[Hughes and Shmatikov 2004] Hughes, D., and Shmatikov, V. 2004. Information hiding, anonymity and privacy: a modular approach. *Journal of Computer Security* 12:3–36.

[Katsuno and Mendelzon 1994] Katsuno, H., and Mendelzon, A. 1994. On the difference between updating a knowledge base and revising it. In Allen, J. F.; Fikes, R.; and Sandewall, E., eds., *KR'91: Principles of Knowledge Representation and Reasoning*. San Mateo, California: Morgan Kaufmann. 387–394.

[Kern-Isberner and Krümpelmann 2011] Kern-Isberner, G., and Krümpelmann, P. 2011. A constructive approach to independent and evidence retaining belief revision by general information sets. In *Proceedings of the 22'nd International Joint Conference on Artificial Intelligence (IJCAI)*.

[Krümpelmann and Kern-Isberner 2008] Krümpelmann, P., and Kern-Isberner, G. 2008. Propagating credibility in answer set programs. In Schwarz, S., ed., *Proc. of the 22nd Workshop on (Constraint) Logic Programming (WLP08), Dresden, Germany*, Technische Berichte. Martin-Luther-Universität Halle-Wittenberg, Germany.

[Lang 2006] Lang, J. 2006. About time, revision and update. In *Proceedings of the 11th Workshop on Nonmonotonic Reasoning (NMR06)*.

[McCullough 1987] McCullough, D. 1987. Specifications for multi-level security and a hook-up property. In *IEEE Symposium on Security and Privacy*, 161–166.

[McLean 1994] McLean, J. 1994. A general theory of composition for trace sets closed under selective interleaving functions. In *Proceedings of the 1994 IEEE Symposium on Security and Privacy*, SP '94, 79–. Washington, DC, USA: IEEE Computer Society.

[Nittka and Booth 2008] Nittka, A., and Booth, R. 2008. A method for reasoning about other agents' beliefs from observations. In Giacomo Bonanno, Wiebe van der Hoek, M. W., ed., *Logic and the Founda- tions of Game and Decision Theory (LOFT 7).*, 153–182. Amsterdam University Press.

[Tamargo et al. 2012] Tamargo, L. H.; García, A. J.; Falappa, M. A.; and Simari, G. R. 2012. Modeling knowledge dynamics in multi-agent systems based on informants. *The Knowledge Engineering Review (KER)* 27(1):84–114.