Framing Faultiness Kripke Style

Hans van Ditmarsch[1] , Krisztina Fruzsa[2,*] , and Roman Kuznets[2,†]

[1] Open University, the Netherlands

[2] TU Wien, Austria
{krisztina.fruzsa,roman.kuznets}@tuwien.ac.at

Epistemic analysis has been used in distributed systems as a potent tool [1,4] for studying agents' uncertainty about the global state of the system, including the global time in asynchronous systems. It is based on the *runs and systems framework* that views global states of a distributed system as possible worlds in a Kripke model. The importance of this methodology is underscored by the broadly applicable *Knowledge of Preconditions Principle* [8], formulated recently by Moses, which states that in all models of distributed systems, if φ is a necessary condition for agent i to perform an action, then agent i knowing that φ holds, written $K_i\varphi$, is also a necessary condition for this agent to perform this action. The agent's complete reliance on its local state as the source of information about the system naturally induces an equivalence relation on the global states, resulting in agents' knowledge being described by the multimodal epistemic logic $\mathsf{S5}_n$.

This epistemic analysis via the runs and systems framework was recently [5,6] extended to *fault-tolerant systems* with so-called *byzantine* agents [7]. (Fully) byzantine agents are the worst-case faulty agents to participate in a distributed system: not only can they arbitrarily deviate from their respective protocols, but their perception of their own actions and the events they observe can be corrupted, possibly unbeknownst to them, resulting in false memories. Whether byzantine agents are actually present in a system or not, the very possibility of their presence has drastic and debilitating effects on the epistemic state of all agents, due to their inability to rule out the so-called *Brain-in-a-Vat Scenario* [9]. In a distributed system, a brain-in-a-vat agent is a faulty agent with completely corrupted perceptions that provide no reliable information about the system [6]. It has been shown that agents' inability to rule out being a brain in a vat precludes them from knowing many basic facts, including their own correctness/faultiness, in both asynchronous [6] and synchronous [10] distributed systems.

The extended runs and systems framework was used in [3] to analyze the *Firing Rebels with Relay* (FRR) problem, a simplified version of the *consistent broadcasting* primitive [11], which has been used as a pivotal building block in distributed algorithms, e.g., for byzantine fault-tolerant clock synchronization, synchronous consensus, etc. Instead of knowledge (unattainable due to the brain-in-a-vat scenario), the analysis of FRR hinges on a weaker epistemic notion called *hope*, which was initially defined as $H_i\varphi := \mathit{correct}_i \to K_i(\mathit{correct}_i \to \varphi)$ and axiomatized in [2] with the help of designated atoms $\mathit{correct}_i$, representing agent i's correctness, as an extension of $\mathsf{K45}_n$ with special axioms regarding atoms $\mathit{correct}_i$.

It turns out that defining faultiness $\mathit{faulty}_i := \neg\mathit{correct}_i$ as inconsistent hopes, i.e.,

$$\mathit{correct}_i \quad := \quad \neg H_i\bot,$$

makes it possible to deal away with designated atoms $\mathit{correct}_i$ and, hence, to avoid the dependency of accessibility relations $\mathcal{H}_i$ for hope modalities H_i on the valuation function in Kripke models for the logic of hope. In this formulation, the logic of hope becomes $\mathsf{KB4}_n$, the logic of the class $\mathcal{KB}4_n$ of transitive and symmetric frames and is axiomatized according to Fig. 1.

*PhD student in the FWF doctoral program LogiCS (W1255).

†Funded by the Austrian Science Fund (FWF) ByzDEL project (P33600).

$$P: \quad \text{all propositional tautologies}$$

$$K^H: \quad H_i(\varphi \to \psi) \wedge H_i\varphi \to H_i\psi \qquad B^H \quad : \quad \varphi \to H_i\neg H_i\neg\varphi$$

$$4^H \ : \quad H_i\varphi \to H_iH_i\varphi$$

$$MP: \quad \frac{\varphi \quad \varphi \to \psi}{\psi} \qquad Nec^H: \quad \frac{\varphi}{H_i\varphi}$$

Figure 1: Axiom system $\mathscr{H}$ for the logic of hope

Theorem 1 (Folklore)**.** *Logic $\mathscr{H}$ is sound and complete with respect to class $\mathcal{KB}4_n$.*

We demonstrate the utility of this reformulation of the logic of hope by encoding a standard limitation on the number of faulty agents in a fault-tolerant distributed system as a *frame-characterizable* property in logic $\mathscr{H}$. It is typical to formulate distributed protocols under the assumption that at most f of the n agents can become faulty ($0 \leq f < n$). This is a natural restriction given that clearly no outcome of agents' protocols can be guaranteed if, e.g., all agents can ignore these protocols. We can encode such requirements by an additional axiom

$$Byz_f \qquad := \qquad \bigvee_{\substack{G\subseteq\mathcal{A}\\|G|=n-f}} \bigwedge_{i\in G} \neg H_i\bot.$$

Remark 2. $Byz_0 = \bigwedge_{i\in\mathcal{A}} \neg H_i\bot$ simply states that all n agents are correct.

Proposition 3. *Axiom Byz_f is characterized by the* all-but-f-seriality *property of frames*

$$(\forall w \in W)(\exists G \subseteq \mathcal{A})\Big(|G| = n - f \wedge (\forall i \in G)\mathcal{H}_i(w) \neq \varnothing\Big),$$

where $\mathcal{H}_i(u) := \{y \in W \mid u\mathcal{H}_i y\}$. In other words, each world must have outgoing arrows for all but f agents.

Definition 4. *Class $\mathcal{KB}4_n^{n-f}$* consists of all frames from $\mathcal{KB}4_n$ that are all-but-f-serial.

Corollary 5. *$\mathscr{H} + Byz_f$ is sound and complete with respect to $\mathcal{KB}4_n^{n-f}$.*

While hope alone is sufficient to restrict the number of faulty agents, we argue that the proper language for reasoning about agents' uncertainty in distributed systems with fully byzantine agents should include both hope H_i and knowledge K_i modalities for all agents. Thus, on the Kripke side, one needs to add accessibility relations $\mathcal{K}_i$ for the K_i modalities. In this language, the connection between knowledge and hope of agent i is represented by the (almost) frame characterizable axiom KH (each direction of equivalence (1) is characterized separately):

$$H_i\varphi \quad \leftrightarrow \quad \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big). \tag{1}$$

Proposition 6. *On the class of frames with shift serial $\mathcal{H}_i$, i.e., with outgoing $\mathcal{H}_i$-arrows whenever there are incoming ones, the right-to-left direction of* (1) *is characterized by frame property $\mathcal{H}\text{in}\mathcal{K}$ stating that $\mathcal{H}_i \subseteq \mathcal{K}_i$.*

Proposition 7. *The left-to-right direction of* (1) *is characterized by frame property* one$\mathcal{H}$ *stating that*

$$(\forall w, v \in W) \qquad \big(\mathcal{H}_i(w) \neq \varnothing \wedge \mathcal{H}_i(v) \neq \varnothing \wedge w\mathcal{K}_i v \quad \Longrightarrow \quad w\mathcal{H}_i v\big).$$

It turns out that the $\mathsf{KB4}_n$ properties of hope can be derived in the combined logic $\mathscr{KH}$ of hope and knowledge that is obtained by extending $\mathsf{S5}_n$ for knowledge modalities with the connection axiom KH from (1) and the *necessary consistency* axiom $H^\dagger := H_i\neg H_i\bot$ for hope ($H^\dagger$ is characterized by shift seriality).

Theorem 8. *Logic $\mathscr{KH}$ is sound and complete with respect to class $\mathcal{KH}$ of models where every $\mathcal{K}_i$ is an equivalence relation, every $\mathcal{H}_i$ is shift serial, and properties $\mathcal{H}$in$\mathcal{K}$ and one$\mathcal{H}$ are satisfied.*

Proposition 9. *In class $\mathcal{KH}$, each accessibility relation $\mathcal{H}_i$ is symmetric and transitive. Hence, $\mathcal{H}_i$ are partial equivalence relations, so that property* one$\mathcal{H}$ *can be described as "no $\mathcal{K}_i$-equivalence class contains more than one $\mathcal{H}_i$-partial-equivalence class."*

Corollary 10 (In fault-free systems, hope is knowledge)**.** $\mathscr{KH} + Byz_0 \quad\vdash\quad H_i\varphi \leftrightarrow K_i\varphi$.

We now use the language of hope and knowledge to formalize the consequences of the *brain-in-a-vat scenario*. These consequences were first established in [6] via a semantic analysis of runs and systems models:

- $iByz := \neg K_i\neg H_i\bot$, i.e., agents cannot reliably establish their own correctness;
- $BiV := H_i\bot \to \neg K_iH_j\bot \wedge \neg K_i\neg H_j\bot$ for $i \neq j$, i.e., a faulty agent lacks any reliable information about other agents, such as whether another agent is correct or faulty.

From these two principles, we can derive by purely syntactic means that no agent knows whether other agents are correct or faulty, as proved in [6] by semantic methods:

Proposition 11. $\mathscr{KH} + iByz + BiV \quad\vdash\quad \neg K_i\neg H_j\bot \wedge \neg K_iH_j\bot \qquad$ *for all $i \neq j$.*

Proposition 12. *Axiom $iByz$ is characterized by the* i-may-aseriality *frame property requiring $(\forall w \in W)\big(\exists w' \in \mathcal{K}_i(w)\big) \;\; \mathcal{H}_i(w') = \varnothing$, stating that each world has a $\mathcal{K}_i$-indistinguishable world with no $\mathcal{H}_i$-outgoing arrows. Axiom BiV for $i \neq j$ is characterized by the* BiValence *frame property requiring*

$$(\forall w \in W) \qquad \Big(\mathcal{H}_i(w) = \varnothing \quad\Longrightarrow\quad \big(\exists w', w'' \in \mathcal{K}_i(w)\big)\big(\mathcal{H}_j(w') \neq \varnothing \wedge \mathcal{H}_j(w'') = \varnothing\big)\Big).$$

We can also easily derive by purely modal means that the brain-in-a-vat scenario is not compatible with fault-free systems: $\mathscr{KH} + Byz_0 \quad\vdash\quad \neg iByz \qquad$ for each $i \in \mathcal{A}$.

Another interesting special case is $f = 1$. On the one hand, half of BiV becomes derivable and, hence, redundant. If any agent and no more than one can be faulty, then agents cannot establish the faultiness of other agents: $\mathscr{KH} + Byz_1 + iByz \quad\vdash\quad \neg K_iH_j\bot \qquad$ for all $i \neq j$.

On the other hand, the other half of BiV leads to undesirable consequences. For $f = 1$, the inability of faulty agents to establish correctness of others would lead to the inability of any agent to establish own faultiness: $\mathscr{KH} + Byz_1 + (H_i\bot \to \neg K_i\neg H_j\bot) \quad\vdash\quad \neg K_iH_i\bot \qquad$ for all $i \neq j$.

Remark 13. Intuitively, if an agent establishes its own faultiness, which does not run afoul of $iByz$ and can be used, e.g., for self-repairing agents, then it will thereby establish the correctness of all other agents. Prohibiting this by the respective half of BiV should be avoided, while the other half is derivable anyway. We, therefore, propose to use $\mathscr{KH} + Byz_f + BiV + iByz$ for $f \geq 2$ or $\mathscr{KH} + Byz_1 + iByz$ for $f = 1$.

Theorem 14. *Axiom system $\mathscr{KHC}$ for common knowledge and common hope consisting of all the axioms of $\mathscr{KH}$ plus the following axioms and inference rules:*

$$Mix^H := C^H_G\varphi \to E^H_G(\varphi \wedge C^H_G\varphi) \qquad Ind^H: \;\; \textit{from } \psi \to E^H_G(\varphi \wedge \psi), \textit{ infer } \psi \to C^H_G\varphi$$
$$Mix^K := C^K_G\varphi \to E^K_G(\varphi \wedge C^K_G\varphi) \qquad Ind^K: \;\; \textit{from } \psi \to E^K_G(\varphi \wedge \psi), \textit{ infer } \psi \to C^K_G\varphi$$

is sound and complete with respect to class $\mathcal{KH}$.

In summary, we provided a description of epistemic views and uncertainties of agents in fault-tolerant distributed systems with fully byzantine agents by means of a multimodal logic with two types of modalities, hope and knowledge (including common hope and common knowledge), proved completeness, and showed how system specifications and properties of such agents can be represented by frame-characterizable properties. This analysis yielded new insights, for instance, into the distinctions between the case of fault-tolerant systems with at most one vs. several byzantine agents. This distinction was already observed in [6] but the newly provided axiomatic representation explains which of the general properties of byzantine agents are violated when all but one agents are correct.

References

[1] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. MIT Press, 1995.

[2] K. Fruzsa. Hope for epistemic reasoning with faulty agents! In *ESSLLI 2019 Student Session*. FOLLI, 2019.

[3] K. Fruzsa, R. Kuznets, and U. Schmid. Fire! In J. Halpern and A. Perea, editors, *Proceedings Eighteenth Conference on Theoretical Aspects of Rationality and Knowledge*, volume 335 of *Electronic Proceedings in Theoretical Computer Science*, pages 139–153. Open Publishing Association, 2021.

[4] J. Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37:549–587, 1990.

[5] R. Kuznets, L. Prosperi, U. Schmid, and K. Fruzsa. Causality and epistemic reasoning in byzantine multi-agent systems. In L. S. Moss, editor, *Proceedings Seventeenth Conference on Theoretical Aspects of Rationality and Knowledge*, volume 297 of *Electronic Proceedings in Theoretical Computer Science*, pages 293–312. Open Publishing Association, 2019.

[6] R. Kuznets, L. Prosperi, U. Schmid, and K. Fruzsa. Epistemic reasoning with byzantine-faulty agents. In A. Herzig and A. Popescu, editors, *Frontiers of Combining Systems, 12th International Symposium, FroCoS 2019, London, UK, September 4–6, 2019, Proceedings*, volume 11715 of *Lecture Notes in Artificial Intelligence*, pages 259–276. Springer, 2019.

[7] L. Lamport, R. Shostak, and M. Pease. The Byzantine Generals Problem. *ACM Transactions on Programming Languages and Systems*, 4:382–401, 1982.

[8] Y. Moses. Relating knowledge and coordinated action: The Knowledge of Preconditions principle. In R. Ramanujam, editor, *Proceedings Fifteenth Conference on Theoretical Aspects of Rationality and Knowledge*, volume 215 of *Electronic Proceedings in Theoretical Computer Science*, pages 231–245. Open Publishing Association, 2016.

[9] A. Pessin and S. Goldberg, editors. *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's the "Meaning of 'Meaning'"*. Routledge, 2015.

[10] T. Schlögl, U. Schmid, and R. Kuznets. The persistence of false memory: Brain in a vat despite perfect clocks. In T. Uchiya, Q. Bai, and I. Marsá Maestre, editors, *PRIMA 2020: Principles and Practice of Multi-Agent Systems: 23rd International Conference, Nagoya, Japan, November 18–20, 2020, Proceedings*, volume 12568 of *Lecture Notes in Artificial Intelligence*, pages 403–411. Springer, 2021.

[11] T. K. Srikanth and S. Toueg. Optimal clock synchronization. *Journal of the ACM*, 34:626–645, 1987.