Definition 1 *The two reward functions* $\mathcal{R}_1 : S^{\otimes} \times 2^{E^{\otimes}} \rightarrow \mathbb{R}$ *and* $\mathcal{R}_2 : S^{\otimes} \times E^{\otimes} \times S^{\otimes} \rightarrow \mathbb{R}$ *are defined as follows.*

$$\mathcal{R}_1(s^{\otimes}, \pi) = \begin{cases} r_{n1}(|E| - |\pi|) & \text{if } [\![s^{\otimes}]\!]_q \notin SinkSet, \\ r_{n1}|E| & \text{if } [\![s^{\otimes}]\!]_q \in SinkSet, \end{cases} \tag{1}$$

where $|E|$ *means number of elements in the set* E *and* r_{n1} *is a negative value.*

$$\mathcal{R}_2(s^{\otimes}, e, s^{\otimes\prime}) = \begin{cases} r_p & \text{if } \exists i \in \{1, \ldots, n\},\ (s^{\otimes}, e, s^{\otimes\prime}) \in \bar{F}_i^{\otimes}, \\ r_{n2} & \text{if } [\![s^{\otimes\prime}]\!]_q \in SinkSet, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

Lemma 1 *For any policy* π *and any recurrent class* $R_{\pi}^{\otimes i}$ *in the Markov chain* $MC_{\pi}^{\otimes}$*,* $MC_{\pi}^{\otimes}$ *satisfies one of the following conditions.*

1. $\delta_{\pi,i}^{\otimes} \cap \bar{F}_j^{\otimes} \neq \emptyset$, $\forall j \in \{1, \ldots, n\}$,
2. $\delta_{\pi,i}^{\otimes} \cap \bar{F}_j^{\otimes} = \emptyset$, $\forall j \in \{1, \ldots, n\}$.

Let $\bar{\mathcal{SV}}_{\varphi}$ be the set of supervisors satisfying the LTL formula φ. For a Markov chain $MC_{SV}^{\otimes}$ induced by a product MDP $D^{\otimes}$ with a supervisor SV, let $S_{SV}^{\otimes} = T_{SV}^{\otimes} \sqcup R_{SV}^{\otimes 1} \sqcup \ldots \sqcup R_{SV}^{\otimes h}$ be the set of states in $MC_{SV}^{\otimes}$, where $T_{SV}^{\otimes}$ is the set of transient states and $R_{SV}^{\otimes i}$ is the recurrent class for each $i \in \{1, \ldots, h\}$, and let $R(MC_{SV}^{\otimes})$ be the union of all recurrent classes in $MC_{SV}^{\otimes}$. Let $\delta_{SV,i}^{\otimes}$ be the set of transtions in a recurrent class $R_{SV}^{\otimes i}$, namely $\delta_{SV,i}^{\otimes} = \{(s^{\otimes}, e, s^{\otimes\prime}) \in \delta^{\otimes}; s^{\otimes} \in R_{SV}^{\otimes i},\ P_T^{\otimes}(s^{\otimes\prime}|s^{\otimes}, e) > 0, P_E^{\otimes}(e|s^{\otimes}, SV(s^{\otimes})) > 0\}$, and let $P_{SV}^{\otimes} : S_{SV}^{\otimes} \times S_{SV}^{\otimes} \rightarrow [0,1]$ such that $P_{SV}^{\otimes} = \sum_{e \in SV(s^{\otimes})} P_T^{\otimes}(s^{\otimes\prime}|s^{\otimes}, e) P_E^{\otimes}(e|s^{\otimes}, SV(s^{\otimes}))$ be the transition probability under SV.

Definition 2 *An accepting recurrent class is defined as the recurrent class whose at least one accepting transition in each accepting set* $\bar{F}_j^{\otimes}$ *with* $j \in \{1, \ldots, n\}$*. We then define the set of index of accepting recurrent classes as* $\mathcal{I}_{Acc}^{SV}$*.*

Theorem 1 *Let* $M^{\otimes}$ *be the product DES corresponding to a DES* M *and an LTL formuula* φ*. Let* $\mathcal{R}_1$ *be a reward function for control patterns. Let* $\bar{k}_{max}$ *denote the maximum expected stopping time of first returning each states in accepting recurrent classes of the Markov chain induced by any supervisor saisfying* φ *as* $\bar{k}_{max} = \max_{SV \in \bar{\mathcal{SV}}_{\varphi}} \max_{i \in \mathcal{I}_{Acc}^{SV}} \{\mathbb{E}^{SV}[k_{s^{\otimes}}^{i,SV} | s_0 = s_{init}^{\otimes}]; s^{\otimes} \in R(MC_{SV}^{\otimes})\}$*, where* $k_{s^{\otimes}}^{i,SV}$ *is the stopping time of first returning to the stste* $s^{\otimes} \in R_{SV}^{\otimes i}$*. If there exists a supervisor* SV *satisfying* φ*, then there exist a discount factor* γ^* *and a positive reward* r_p^* *that satisfies* $\gamma^{\bar{k}_{max}} r_p^* > (1 + \ldots + \gamma^{\bar{k}_{max}})||\mathcal{R}_1||_{\infty}$ *such that any algorithm that maximizes the expected discounted reward with* $\gamma > \gamma^*$ *and* $r_p > r_p^*$ *will find a supervisor satisfying* φ*.*

Proof 1 *Suppose that SV^* be an optimal supervisor but does not satisfy the LTL formula φ or there is a state $s^{\otimes}_{sink}$ reachable from the initial state such that $[\![s^{\otimes}_{sink}]\!]_q \in SinkSet$ under the supervisor SV^*. Then, for any recurrent class $R^{\otimes i}_{SV^*}$ in the Markov chain $MC^{\otimes}_{SV^*}$ and any accepting set $\bar{F}^{\otimes}_j$ of the product DES $M^{\otimes}$, $\delta^{\otimes}_{SV^*,i} \cap \bar{F}^{\otimes}_j = \emptyset$ holds for the first case by Lemma 1 and there is a recurrent class $R^{\otimes i}_{SV^*}$ such that $s^{\otimes}_{sink} \in R^{\otimes i}_{SV^*}$ for the second case. We consider the two cases separately.*

1. *Assume that SV^* does not the LTL formula φ. By the assumption, the system under the supervisor SV^* can obtain rewards only in the set of transient states. We consider the best scenario in the assumption. Let $p^k(s,s')$ be the probability of going to a state s' in k time steps after leaving the state s, and let $Post(T^{\otimes}_{\pi^*})$ be the set of states in recurrent classes that can be transitioned from states in $T^{\otimes}_{\pi^*}$ by one event occurrence. For the initial state $s^{\otimes}_{init}$ in the set of transient states, it holds that*

$$
\begin{aligned}
V^{SV^*}(s^{\otimes}_{init}) = & \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{\pi^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) \\
& \sum_{s^{\otimes\prime} \in T^{\otimes}_{\pi^*} \cup Post(T^{\otimes}_{\pi^*})} \sum_{e \in SV(s^{\otimes})} P^{\otimes}_T(s^{\otimes\prime}|s^{\otimes}, e) P^{\otimes}_E(e|s^{\otimes}, SV(s^{\otimes})) \mathcal{R}(s^{\otimes}, SV(s^{\otimes}), e, s^{\otimes\prime}) \\
\leq & \; r_p \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{\pi^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}).
\end{aligned}
$$

   By the property of the transient states, for any state $s^{\otimes}$ in $T^{\otimes}_{\pi^*}$, there exists a bounded positive value m such that $\sum_{k=0}^{\infty} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) \leq \sum_{k=0}^{\infty} p^k(s^{\otimes}_{init}, s^{\otimes}) < m$ [1]. Therefore, there exists a bounded positive value $\bar{m}$ such that $V^{\pi^*}(s^{\otimes}_{init}) < \bar{m}$.

2. *Assume that there is a state $s^{\otimes}_{sink}$ reachable from the initial state such that $[\![s^{\otimes}_{sink}]\!]_q \in SinkSet$ under SV^*. By the assumption, there is at least one recurrent class $R^{\otimes i}_{SV^*}$ reachable from the initial state such that $s^{\otimes}_{sink} \in R^{\otimes i}_{SV^*}$. We consider the best scenario in the assumption.*

$$
\begin{aligned}
V^{SV^*}(s^{\otimes}_{init}) < & Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi) \sum_{k=0}^{\infty} \gamma^k (r_p + ||\mathcal{R}_1||_{\infty}) + \gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink}) \sum_{k=0}^{\infty} \gamma^k r_{n2} \\
= & \frac{1}{1-\gamma} \{ Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi)(r_p + ||\mathcal{R}_1||_{\infty}) + \gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink}) r_{n2} \}.
\end{aligned}
$$

   Therefore, if it holds that $r_{n2} < -\frac{Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi)}{\gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink})}(r_p + ||\mathcal{R}_1||_{\infty})$, we have $V^{SV^*}(s^{\otimes}_{init}) < 0$ for any $\gamma \in [0,1)$.

Let $\bar{SV}$ be a supervisor satisfying φ. We consider the following two cases.

1. *Assume that the initial state $s^{\otimes}_{init}$ is in a recurrent class $R^{\otimes i}_{\bar{\pi}}$ for some $i \in \{1, \ldots, h\}$. For any accepting set $\bar{F}^{\otimes}_j$, $\delta^{\otimes}_{\bar{\pi},i} \cap \bar{F}^{\otimes}_j \neq \emptyset$ holds by the definition of $\bar{\pi}$. The expected discounted reward for $s^{\otimes}_{init}$ is given by*

$$V^{\bar{SV}}(s^{\otimes}_{init}) = \mathbb{E}^{SV}[\sum_{k=0}^{\infty}\gamma^k \mathcal{R}(s_k, \pi_k, e_k, s_{k+1})|s_0 = s^{\otimes}_{init}] \tag{3}$$

Since $s^{\otimes}_{init}$ is in $R^{\otimes i}_{\bar{\pi}}$, there exists a set of positive numbers $K = \{k \ ; \ k \geq n, p^k(s^{\otimes}_{init}, s^{\otimes}_{init}) > 0\}$ [1]. We consider the worst scenario of returning the initial state in this case. For the stopping time k of first returning to the initial state, it holds that

$$\begin{aligned} V^{\bar{\pi}}(s^{\otimes}_{init}) >& \mathbb{E}^{\bar{SV}}[\gamma^k r_p - (1 + \ldots + \gamma^k))||\mathcal{R}_1||_{\infty} + \gamma^k V^{\bar{\pi}}(s^{\otimes}_{init})|s_0 = s^{\otimes}_{init}] \\ \geq& \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]} r_p - (1 + \ldots + \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]})||\mathcal{R}_1||_{\infty} + \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]} V^{\bar{\pi}}(s^{\otimes}_{init}) \\ =& \frac{\gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]} r_p - (1 + \ldots + \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]})||\mathcal{R}_1||_{\infty}}{1 - \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]}}, \end{aligned}$$

where second inequality holds since it holds that $\mathbb{E}^{\bar{SV}}[\gamma^k|s_0 = s^{\otimes}_{init}] \geq \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]}$ and $\frac{1-\gamma^{\mathbb{E}^{\bar{SV}}[k+1|s_0=s^{\otimes}_{init}]}}{1-\gamma} \leq \mathbb{E}^{\bar{SV}}[\frac{1-\gamma^{k+1}}{1-\gamma}|s_0 = s^{\otimes}_{init}]$ by Jensen's inequality. Therefore, for any $\bar{m} \in (V^{SV^*}(s^{\otimes}_{init}), \infty)$ and any reward function $\mathcal{R}_1$ defined by definition 1, there exist $\gamma^* < 1$ and a positive reward r^*_p that satisfies $\gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]} r^*_p > (1 + \ldots + \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]})||\mathcal{R}_1||_{\infty}$ such that $\gamma > \gamma^*$ and $r_p > r^*_p$ imply $V^{\bar{SV}}(s^{\otimes}_{init}) > \bar{m} > V^{SV^*}(s^{\otimes}_{init})$.

2. *Assume that the initial state $s^{\otimes}_{init}$ is in the set of transient states $T^{\otimes}_{\bar{SV}}.P^{M^{\otimes}}_{\bar{SV}}(s^{\otimes}_{init} \models \varphi) > 0$ holds by the definition of $\bar{SV}$. For a recurrent class $R^{\otimes i}_{\bar{SV}}$ such that $\delta^{\otimes}_{\bar{SV},i} \cap \bar{F}^{\otimes}_j \neq \emptyset$ for each accepting set $\bar{F}^{\otimes}_j$, there exist a number $\bar{l} > 0$, a state $\hat{s}^{\otimes}$ in $Post(T^{\otimes}_{\bar{SV}}) \cap R^{\otimes i}_{\bar{SV}}$, and a subset of transient states $\{s^{\otimes}_1, \ldots, s^{\otimes}_{\bar{l}-1}\} \subset T^{\otimes}_{\bar{SV}}$ such that $p(s^{\otimes}_{init}, s^{\otimes}_1) > 0$, $p(s^{\otimes}_i, s^{\otimes}_{i+1}) > 0$ for $i \in \{1, ..., \bar{l} - 2\}$, and $p(s^{\otimes}_{\bar{l}-1}, \hat{s}^{\otimes}) > 0$ by the property of transient states. Hence, it holds that $p^{\bar{l}}(s^{\otimes}_{init}, \hat{s}^{\otimes}) > 0$ for the state $\hat{s}^{\otimes}$. Thus, for the stopping time k of first returning to the state $\hat{s}^{\otimes}$, by ignoring positive rewards in $T^{\otimes}_{\bar{\pi}}$ and assuming the system incurs the full costs with regard to disabling events, we have*

$$
\begin{aligned}
&V^{\bar{SV}}(s^{\otimes}_{init})\\
=&\mathbb{E}^{SV}[\sum_{m=0}^{\infty}\gamma^m\mathcal{R}(s_m,\bar{SV}(s_m),e_m,s_{m+1})|s_0=s^{\otimes}_{init}]\\
\geq&\mathbb{E}^{SV}[-\sum_{m=0}^{l'}\gamma^m||\mathcal{R}_1||_{\infty}|s_0=s^{\otimes}_{init}]\\
&+\mathbb{E}^{SV}[\gamma^{l'}\sum_{m=0}^{\infty}\gamma^m\mathcal{R}(s_{m+l'},\bar{SV}(s_{m+l'}),e_{m+l'},s_{m+l'+1})|s_0=s^{\otimes}_{init}]\\
>&\gamma^l p^l(s^{\otimes}_{init},\hat{s}^{\otimes})\mathbb{E}^{\bar{SV}}[\gamma^k r_p-(1+\ldots+\gamma^k)||\mathcal{R}_1||_{\infty}+\gamma^k V^{\bar{SV}}(\hat{s}^{\otimes})|s_l=\hat{s}^{\otimes}]-\frac{1-\gamma^{\bar{l}}}{1-\gamma}||\mathcal{R}_1||_{\infty}\\
\geq&\gamma^l p^l(s^{\otimes}_{init},\hat{s}^{\otimes})\\
&\{\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]}r_p-(1+\ldots+\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]})||\mathcal{R}_1||_{\infty}+\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]}V^{\bar{SV}}(\hat{s}^{\otimes})\}-\frac{1-\gamma^{\bar{l}}}{1-\gamma}||\mathcal{R}_1||_{\infty}\\
=&\gamma^l p^l(s^{\otimes}_{init},\hat{s}^{\otimes})\\
&\frac{\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]}r_p-(1+\ldots+\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]})||\mathcal{R}_1||_{\infty}}{1-\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]}}-\frac{1-\gamma^{\bar{l}}}{1-\gamma}||\mathcal{R}_1||_{\infty},
\end{aligned}
$$

where $\bar{l}=\mathbb{E}^{\bar{SV}}[l'|p^{l'}(s^{\otimes}_{init},\hat{s}^{\otimes})>0]$. *Therefore, for any* $\bar{m}\in(V^{SV^*}(s^{\otimes}_{init}),\infty)$ *and any reward function* $\mathcal{R}_1$ *defined by definition 1, there exist* $\gamma^*<1$ *and a positive reward* r_p^* *that satisfies* $\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]}r_p^*>(1+\ldots+\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]})||\mathcal{R}_1||_{\infty}$ *such that* $\gamma>\gamma^*$ *and* $r_p>r_p^*$ *imply* $V^{\bar{SV}}(s^{\otimes}_{init})>\bar{m}>V^{SV^*}(s^{\otimes}_{init})$.

The results contradict the optimality assumption of SV^*

References

[1] R. Durrett, *Essentials of Stochastic Processes*, 2nd Edition. ser. Springer texts in statistics. New York; London; Springer, 2012.

[2] L. Breuer, “Introduction to Stochastic Processes,” [Online]. Available: https://www.kent.ac.uk/smsas/personal/lb209/files/sp07.pdf

[3] S.M. Ross, *Stochastic Processes*, 2nd Edition. University of California, Wiley, 1995.

[4] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesv́ari, “Convergence results for single-step on-policy reinforcement learning algorithms” *Machine Learning*, vol. 38, no. 3, pp, 287–308, 1998.

[5] J. Kretínský, T. Meggendorfer, S. Sickert, "Owl: A library for ω-words, automata, and LTL," in *Proc. 16th International Symposium on Automated Technology for Verification and Analysis*, 2018, pp. 543–550.