Reinforcement Learning for Electric Power System Decision and Control: Past Considerations and Perspectives

Mevludin Glavic * **Raphaël Fonteneau** * **Damien Ernst** *

* *Dept. of Electrical Engineering and Computer Science, University of Liège, Allée de la découverte 10, 4000 Liège, Belgium (e-mail: raphael.fonteneau@ulg.ac.be)*

Abstract:
In this paper, we review past (including very recent) research considerations in using reinforcement learning (RL) to solve electric power system decision and control problems. The RL considerations are reviewed in terms of specific electric power system problems, type of control and RL method used. We also provide observations about past considerations based on a comprehensive review of available publications. The review reveals the RL is considered as viable solutions to many decision and control problems across different time scales and electric power system states. Furthermore, we analyse the perspectives of RL approaches in light of the emergence of new-generation, communications, and instrumentation technologies currently in use, or available for future use, in power systems. The perspectives are also analysed in terms of recent breakthroughs in RL algorithms (Safe RL, Deep RL and path integral control for RL) and other, not previously considered, problems for RL considerations (most notably restorative, emergency controls together with so-called system integrity protection schemes, fusion with existing robust controls, and combining preventive and emergency control).

Keywords: Electric power system, reinforcement learning, control, decision.

1. INTRODUCTION

Societal and economic costs of large electric power systems' blackouts could be as high as 10 billion dollars with 50 million people affected, as estimated for the US-Canada Power System Outage of August 14, 2003 US-DoE (2004). The blackouts are often caused by poor design and coordination of controls needed to operate electric power systems. Thus, in order to prevent (or at least decrease) the probability of large blackouts, research, development and implementation of improved controls are of paramount importance.

Electric power system control techniques have always evolved in order to fulfil the requirements of the electric power industry during its development. As usually happens, the development of new control techniques in power systems is often based on advances achieved in applied mathematics, control theory, computer science, operational research, telecommunications, and the availability of more powerful computational facilities.

Increasing complexity of todays' large interconnected power systems (including uncertainties brought about by encouragement of the use of so-called renewable energy sources) requires advanced control techniques to be applied to control power systems more efficiently.

Better communications' infrastructure, stronger computational capabilities, and new control devices (such as power electronics) open up the possibilities to implement advanced control schemes which are able to process the observations realized on the power system and to control it appropriately. Embedding the learning methods into the control schemes is an effective way to endow the controllers with the capability to learn and update their decision-making.

RL offers a panel of methods that allows controllers to learn a goal-oriented control law from interactions with a system or its simulation model Sutton and Barto (1998); Busoniu et al. (2010). RL driven controllers (agents) observe the system state, take actions, and observe the effects of these actions. They process the accumulated experience and progressively learn an appropriate control law, i.e. an algorithm to associate suitable actions to the observations in order to fulfil a pre-specified objective Sutton and Barto (1998); Busoniu et al. (2010).

The potentials of RL to solve electric power system control and decision problems was recognized by the research community and a number of considerations were offered. This paper surveys past and recent RL considerations to solve power system control and decision problems and suggest some future research direction.

To the best of our knowledge there is no document like this and we hope it may serve as a reference document for future considerations of RL in electric power system decision and control problems.

The paper is organized as follows. Section 2 briefly introduces electric power system control and decision problems. In Section 3, a framework for RL considerations in electric power systems is presented. Section 4 surveys past and more-recent research publications together with major observations made by the authors. Section 5 proposes some future research directions, while section 6 concludes.

2. ELECTRIC POWER SYSTEM DECISION AND CONTROL: A SHORT OVERVIEW

Electric power systems face a multitude of control problems over different operating states and time scales. The widely accepted classification of electric power system operating states is the one introduced in DyLiacco (1974) Fig. 1 illustrates five operating states as defined in DyLiacco (1974) and adapted in Padiyar (2008).

The states are defined in terms of the status of equality (E) and inequality (I) constraints of the system (violated (indicated with "~" in Fig. 1) or not violated). The equality constraints express the generation-load demand balance, while inequality constraints express physical limitations of power system components (usually defined in terms of current and voltage magnitudes, active, reactive and apparent powers that a system component can withstand without any damage).

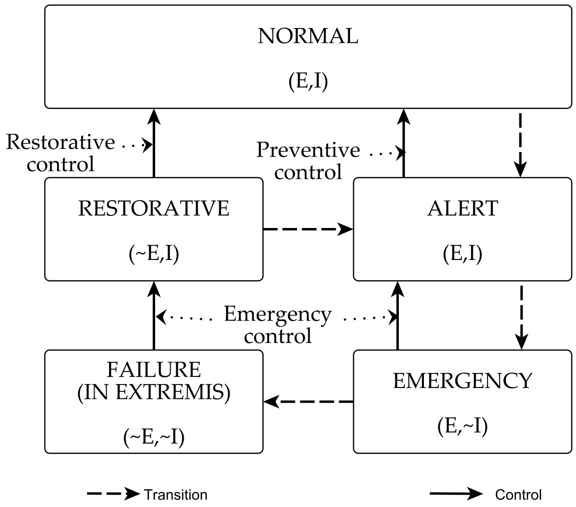

Fig. 1. Power system operating states Padiyar (2008)

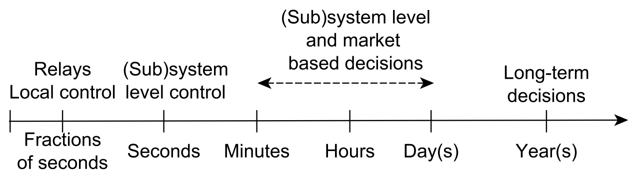

Fig. 2. Power system decision and control time scale decomposition

Fig. 2 depicts power system decisions and controls ranging from fractions of seconds to year(s). Time decomposition shown in Fig. 2 corresponds to the normal and alert system states.

Examples of fast controls, acting in fractions of seconds, are protective relays, automatic voltage regulators on generators, turbine governors, etc., while (sub)system level controls, acting over several seconds, include for example automatic generation control, secondary voltage control, etc. (Sub)system level controls, acting over minutes, hours or a day include scheduling problems such as economic dispatch.

Market-based decisions also relate to scheduling, while longer-term decisions include maintenance scheduling or the use of hydro potentials. Other types of controls, such as preventive and emergency controls (see Fig. 1), are usually designed as discrete controls activated upon detection of the system constraints' violations.

In the next section we review past and more-recent considerations of RL to solve different types of electric power system control/decision problems.

3. A FRAMEWORK FOR RL CONSIDERATION IN POWER SYSTEM DECISION AND CONTROL

The RL controller (agent) interacts with an environment (system) by observing states and selecting actions. After each moment of interaction, the agent receives a feedback signal, reinforcement signal, or reward from the system being delivered to the learning system in response to the execution of control action Sutton and Barto, 1998; Busoniu et al., 2010. The most commonly studied objective is to maximize, for each time step, the expected sum of future reinforcements or discounted return defined as the sum of rewards over future time steps Sutton and Barto, 1998; Busoniu et al., 2010.

A likely framework for application of the RL in power system decision and control is illustrated in Fig. 3. It defines two modules: learning and execution. The learning module is a typical RL implementation, while the execution module is a simple greedy agent that uses the knowledge gained in the learning module to make controls/decisions.

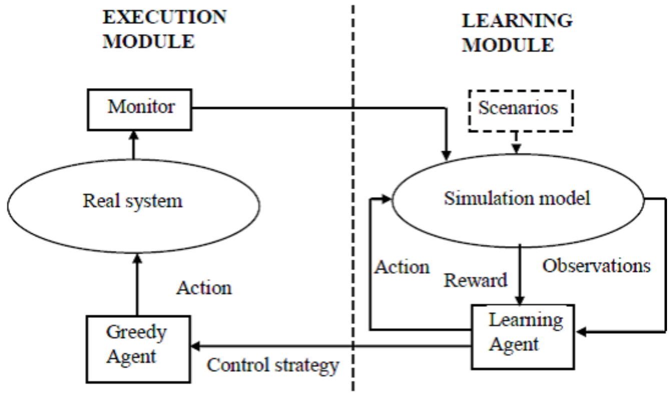

Fig. 3. A likely framework for RL in power system decision and control

The RL learning agent interacts with the system model (interaction with a real system is a concern since exploration in the search for an optimal policy in RL could create unsafe and unreliable system situations). The periodical transfer of a learned strategy to a greedy agent ensures improvement in time.

Note also that the scenarios used for the learning agent are enriched by events observed in a real system (for their replication) since it is not always possible to capture all possible scenarios in a simulation model. In this way RL decision/control benefits from both a simulation environment and a real system while direct interaction with the system is avoided.

4. PAST AND RECENT CONSIDERATIONS OF RL FOR ELECTRIC POWER SYSTEM DECISION AND CONTROL

In this section we survey past and recent considerations through a summary in tabular form and main observations of the authors. A large number of publications on this problem are available. However, we are only focusing on the major ones (those published in journals).

Table 1. summarizes considerations in terms of power system control/decision problem, type of of control and RL method used.

Note that if the cells in columns of Table 1. are left blank then they are assumed to be the same as first filled cell above.

Instead of providing short descriptions of every consideration listed in Table 1., we provide our observations based on a comprehensive review of publications mentioned in Table 1.

We offer the following observations:

- RL was considered in a variety of electric power system control and decision problems (listed in Table 1). This confirms the potential of RL methods and the interest of power system researchers/engineers in using these methods to solve problems.
- RL was considered as control/decision across all over system operating states shown in Fig. 1 and across different time scales shown in Fig. 2 (note that normal control in Table 1 means the controls are in a normal system operating state but also refer to decisions like scheduling and dispatch).
- All past and recent considerations are based on off-line RL, i.e. good control policies are learned from a system model. This is understandable for, at least, two reasons: the first is the designers of power system controls have been always concerned first and foremost with safety and reliability, and only techniques that are fully understood and which are guaranteed to work, are used, and the second is the exploration in RL would be dangerous if direct interaction between control and the system is in place.
- Prevailing RL methods used are Q-learning and some of its variants, such as $Q(\lambda)$ and fitted Q iteration. This prevalence is also explained by the need to use fully understood methods (at least fully understood by designers of power system controls and decisions). Other considerations include $R(\lambda)$, temporal-difference (TD) and policy search.
- Domain-specific knowledge, or the knowledge about the particular problem under study, plays an important role in making problem solving tractable. The usual roles of domain knowledge are:
  - · Making the computations necessary for solving the problem more time- or space-efficient,
  - · Guiding the solution process,
  - · Increasing robustness (especially when unforeseen or previously un-experienced situations occur), and
  - · Allowing a learning agent to perform at a satisfactory level even at the beginning of learning.

Table 1. Summary of RL considerations for electric power system control/decision

Problem	Type of control	RL method	Reference(s)
Electricity market simulation	Market decision	Q-learning	Harp et al. (2000) Rahimiyan et al. (2010) Nanduri and Das (2007) Lincoln et al. (2012) Kim et al. (2016) Krause et al. (2006)
Transient angle instability	Emergency	Q-learning	Ernst et al. (2004) Glavic (2005) Glavic et al. (2005a) Glavic et al. (2005b) Li and Wu (1999)
		Fitted Q iteration	Ernst et al. (2009)
		Policy search	Mohagheghi et al. (2006)
Oscillatory angle instability	Emergency	Q-learning	Ernst et al. (2004) Wang et al. (2014) Glavic et al. (2005a) Ademoye and Feliachi (2012) Karimi et al. (2009)
Voltage control	Normal	Q-learning	Xu et al. (2012) Vlachogiannis et al. (2004)
AGC (Automatic) generation control)	Normal	$Q(\lambda)$ with elig. traces	Yu et al. (2011)
		Q-learning	Daneshfar and Bevrani (2010) Ahamed et al. (2002)
		$R(\lambda)$	Yu et al. (2012b)
Economic dispatch	Normal	Q-learning	Jasmin et al. (2011) Yu et al. (2016)
Wide-area control	Emergency	TD Q-learning	Yousefian et al. (2016) Yan et al. (2016) Hadidi and Jeyasurya (2013)
Households control	Normal	Q-learning	Wang et al. (2016) Yan et al. (2016)
Wind generation control	Normal	Q-learning	Wei et al. (2015) Tang et al. (2015)
		$Q(\lambda)$	Yu et al. (2012a)
Demand control	Normal	Fitted Q iteration	Ruelens et al. (2016) Vandael et al. (2015)
System restoration	Restorative	Q-learning	Ye et al. (2011)
Congestion management	Emergency	Q-learning	Zarabbian et al. (2016)
Microgrids control	Normal	Q-learning	Khorramabady et al. (2015) Li et al. (2012)
		Policy search	Venayagamorthy et al. (2016)

One machine infinite bus (OMIB) transformation was used as the domain-specific knowledge for the problem dimensionality reduction in most of the RL considerations for transient and oscillatory angle instability problems (see for example: Ernst et al. (2004); Glavic (2005); Glavic et al. (2005a); Ernst et al. (2009)). Work presented in Ruelens et al. (2016) extended the fitted Q iteration algorithm in order to take advantage of domain-specific knowledge (in particular case a forecast of the exogenous data is provided to design demand response control). $Q(\lambda)$ with eligibility traces is used to take advantage of domain-specific knowledge in Yu et al. (2011).

- Even with the use of domain-specific knowledge, the dimensionality of the problem often becomes an issue since, in many cases, the problem is revealed to be partially observable Markov decision problems and the use of history of inputs is needed to recover Markov property.
- In general, there is a lack of efficient fusion of RL models with control theory and practice for power systems. The work presented in Glavic et al. (2005a) suggested fusion of RL and the concept of control Lyapunov functions (this approach is further elaborated in Glavic et al. (2006)). Other works suggesting fusion of RL with known control techniques include Ernst et al. (2009); Wang et al. (2014) where RL was considered together with model predictive control, and Li and Wu (1999) where RL is combined with fuzzy logic control.
- RL was considered as a single agent (where RL controls individual power system components) or a multi-agent system for the problem solution. Power system components considered include: dynamic brake Ernst et al. (2004); Glavic (2005), thyristor controlled series capacitor Ernst et al. (2004, 2009), quadrature booster Li and Wu (1999), synchronous generators (all AGC related references), individual or aggregated loads Vandael et al. (2015); Ruelens et al. (2016), etc. If used as a multi-agent system, then additional state variables must be introduced to ensure convergence of these essentially distributed computation schemes, and an adapted variant of standard RL methods is often used (for example correlated equilibrium $Q(\lambda)$ Yu et al. (2012a)).

In the remainder of this paper we provide some research and development opportunities in terms of recent breakthroughs in RL together with the control and decision challenges in future electric power systems and perspectives in restorative, emergency, robust, and distributed control.

5. PERSPECTIVES OF RL FOR ELECTRIC POWER SYSTEM DECISION AND CONTROL

Although a number of power system control/decision problems were considered with application of RL methods, we believe these methods could be extended to a variety of other problems.

In this section we provide our view on the RL perspectives for electric power systems decision and control considerations. The perspectives are analysed in terms of new and more efficient RL methods and future needs of developing power systems where even more decision and control is obviously needed due to structural changes and increased uncertainties in the system operation.

5.1 Recent Breakthroughs in the RL Community

The breakthroughs relevant for electric power system decision and control considerations are Safe RL (for allowing on-line interaction with the system), Deep RL (for better handling of data sets) and path integral control approach to RL (for higher efficiency and fewer open parameters in RL).

Safe RL Designers of power system controls have always been concerned first and foremost with safety and reliability. In this respect new algorithmic solutions coming in the RL field as safe RL are worthy of considerations for revisiting some past applications of RL to power system control problems, as well as exploring new problems of electric power system decision and control.

Most important results in this field are those presented in Thomas (2015). Relevant results we found in Thomas (2015) are those related to the improvements of fitted Q iteration and least squares policy iterations algorithms. Proposed algorithms balance the trade-off between predicted performance and predicted lower bound when searching for safe policies, i.e. policies that are guaranteed to improve upon a user-specified baseline with a user-specified confidence level (Thomas (2015)).

This could open the possibility of using RL methods in on-line mode where an RL-driven agent interacts directly with the system since safe RL searches for new and improved policies, while ensuring that the probability of a harmful policy is low (Thomas (2015)). At the same time, the user specifies the meaning of harmful and how low the probability should be in order to define the level of acceptable risk (Thomas (2015)) through the domain-specific knowledge. This is particularly important in cases where the phenomena are difficult to model and reproduce within a simulation environment. An excellent source on safe RL is the work presented in Garcia and Fernandez (2015).

We believe that the prevailing approach is going to be a combined implementation of RL (off-line and on-line) where it first learns using a system model then it further improves its behaviour through interaction with the real system (safe RL will still be needed for the on-line interaction).

Note that safe (on-line) RL may also be combined with batch RL, a subfield of RL for which it is assumed that the sole knowledge available about the problem is a set of interactions with the real system embedded in batch of trajectories. In such a setting, safety is related with the generalization properties of the batch RL algorithm used to process the batch of trajectories. Designing batch RL algorithms with cautious generalization properties has been addressed using synthesized trajectories in Fonteneau et al. (2013b), and also operation research techniques in Fonteneau et al. (2013a).

Deep RL Recent years have shown huge progress in RL algorithms for control and decision making problems with large state spaces (typically, matrices of pixels). This performance increase is mainly due to the import of deep learning techniques that have recently become very mature (see Bengio (2009); LeCun et al. (2015); Schmidhuber (2015)).

These techniques allow for the learning of rich features from massive sets of data, in particular sets (or time-series) of images. This automatic feature extraction ability has successfully be incorporated to RL paradigms, allowing classic RL techniques such as ϵ-greedy Q-Learning (Watkins and Dayan (1992)) attain state-of-the-art performance when playing the game of Go (Silver et al. (2016)) or Atari video games (Mnih et al. (2015)). This has opened the so-called Deep RL area.

Encouraged by these successes, the Deep RL field of research is revisiting some of the main works achieved in the past three decades using deep learning function approximators (see Busoniu et al. (2010) for a global view of the use of function approximators in RL). In particular, the deep learning remastering of Double Q Learning (see Hasselt et al. (2015) and Hasselt (2010) for the original version without deep learning) or Memory Replay Schaul et al. (2015) ideas shows that RL promises that were originally proposed decades ago are definitely worth revisiting in the light of Deep Learning architecture.

These recent Deep RL successes should naturally be seen as invitations to revisit RL-based optimal control approaches that have been proposed to address power system problems in the past, as well as the ones with prospective applications in the future.

Deep RL could also be combined with safety-based controls offering a safe deep RL for safety-critical applications, such as the majority of power system control/decision problems.

Path integral control approach to RL This is an example of the combination of classical techniques from optimal control and dynamic programming with modern learning techniques from statistical estimation theory.

Theodorou et al. (2010) suggest the use of the framework of stochastic optimal control with path integrals to derive a novel approach to RL with parameterized policies. In principle, the policy improvement are achieved by its transformation into an approximation problem of a path integral with the exploration noise as the only open algorithmic parameter. The algorithm can be realized as either model-based, semi-model-based, or model free, depending on how the learning problem is structured. More precisely, Theodorou et al. (2010) introduced policy improvement with the path integrals PI^2 method.

We believe that such a combination of classical technique from optimal control and RL could offer efficient and numerically robust RL methods to be used in power system decision and control problems.

5.2 The future of electric power systems: even more control and decision-making opportunities

In this section, we propose a prospective view of the main trends that will follow electric power systems in the coming years, or perhaps decades. We also provide a few research directions involving reinforcement learning.

Main trends The future of the electric power system may be characterized by the following trends (Ekanayake et al. (2012)):

- A progressive deployment of distributed electricity generation capacities, associated with a growing deployment of distributed storage capacities, and thus inducing a deployment of electric microgrids,
- The rise of digitalization, inducing an increase of available data, pushing the whole electric power system into the “smart grid era”,
- The rise of a two-sided market-type economy, mainly driven by the development of apps on mobile devices,
- The introduction of blockchains, distributed ledgers and smart-contracts, and more generally, the possibility to automate and distribute transactions,
- The progressive emergence of electric vehicles (EVs), that may also be seen as mobile storage capacity.

At a first glance, the previous list of items may introduce additional complexity in the traditional operation of electrical power systems. The previously described trends will definitely lead to the introduction of additional control variables in the global energy landscape. However, many interesting opportunities may also be brought about by such control variables, and RL (particularly Safe and Deep RL) could offer practical solutions to these problems.

Towards decentralized electricity generation, distributed storage and electricity microgrids The last decade has seen the rapid emergence of renewable electricity production capacities, in particular photovoltaic (PV) panels and wind turbines. These renewable electricity generation techniques have the characteristic to be strongly “decentralizable”, i.e. to be installed at a rather small scale (typically a few kW power), and close to electricity consumers.

In addition, there is growing evidence that electricity storage technologies emerge as the key enablers of future power system operation Ekanayake et al., 2012.

As a consequence of the last two above-mentioned points, electrical microgrids may progressively become economically viable in the coming years, depending on the geographical situation, on the possibility to install PV panels or small wind turbines, and on the load profile of the (local) consumer(s). More than 160 microgrid project are currently active world-wide (see http://www.resilient-project.eu/). A growing research community is investigating business cases related with electricity microgrids François-Lavet et al. (2016a).

Very recent works have already proposed the application of machine learning techniques such as imitative learning Aittahar et al. (2015) and Deep RL techniques François-Lavet et al. (2016b) for planning under uncertainty within electricity microgrids. It appears that favourable control architecture of microgrids resemble the hierarchical control of traditional power systems. This suggests previous achievements (including RL) could be appropriately adopted and re-used. Also, a RL hierarchical model proposed in Dalal et al. (2016), where the goal is set to be electrical grid reliability maximization with consideration of more than a single stakeholder in system operation, is worth of further exploration in this context.

Several research investigations remain to be carried out, in particular those related to the interaction between microgrids and the grid, and also among microgrids. Here we also see opportunities for RL considerations.

The smart grid era: towards more and more data-driven solutions Even if "smart grid" is a term that remains fuzzy in its definition, one main feature that characterizes the smart grid era is the availability and the exploitation of more and more data, mainly acquired from new generation smart meters and other sensors.

Opportunities offered by data analytics are manifold, especially in the context of the rise of renewable (and intermittent) sources of electricity. Perhaps one of the most promising research direction opened in this field is Active Network Management (ANM), and the possibility to adapt the load based on certain production scenarios. For instance, an ANM benchmark based on the formalization of the ANM problem as a Markov decision process was developed in Gemine et al. (2014). This ANM benchmark offers an opportunity to put recent Deep RL techniques to the service of electric power systems.

Two-sided markets and sharing economy Two-sided markets were probably first formalized in the context of the popularization of payment cards Rochet and Tirole (2003). On the one side, customers were interested in getting a payment card offering a wide range of places where it could be used, and on the other side, sellers were interested in offering payment facilities to payment cards offering a large pool of customers.

The notion of two-sided market economy may become central in electric power systems in the coming years, mainly because of the distributed nature of renewable energy production capacities. Electricity consumers have progressively become electricity prosumers. In the future, it may not be impossible that energy producers and consumers exchange energy using ad hoc platforms. These exchange platforms should probably emerge rapidly as soon as electric vehicles increase their market share.

Electric vehicles and storage: challenges and opportunities Electric Vehicles (EVs) have recently revealed huge improvements in terms of autonomy and pricing. EVs' load may represent a consequent additional consumption of electricity. Uncertainties in charging patterns further complicates operation and control of the system. At the same time EVs offer an great opportunity to manage the fluctuations of the production of electricity using renewable energy Kydd et al. (2016).

New opportunities are opening for EVs (and in general to other storage capable devices) in fast system frequency regulation. Many electric power utilities across the USA already use, or are in the stage of implementation, so-called performance based frequency regulation (needed in case of any imbalance between electricity generation and demand). EVs could also be "controlled" to select their charging stations and their time of charge to avoid electrical overvoltages (Olivier et al. (2016)) or congestions and favor the integration of renewable energy into distribution network (Dubois et al. (2017)).

The aim is to encourage power system devices able to provide fast response to participate in the regulation and financial incentives are put in place for their performances (faster frequency stabilization then uses traditional approaches based on big but slow generation units) Glavic and Alvarado (2016). If participating in these fast regulation schemes, EVs (if aggregated in enough number to provide this service) could be controlled as a dynamic brake using RL as demonstrated in Glavic (2005) and the same holds true for any other storage device mentioned earlier.

Blockchain and distributed ledgers The emergence of electricity prosumers will also push for an evolution of the "traditional transaction structure" currently linking all agents producing, transporting, distributing and consuming electricity.

Recently, several projects have emerged around the notion of electricity prosumers sharing their electricity while managing transactions using blockchain and smart contract technologies. This is, for instance, the case of the Brooklyn Microgrid project [1].

A distributed ledger (also called shared ledger) is a consensus of replicated, shared, and synchronized digital data geographically spread across multiple sites [2]. Certainly, it offers a transparent, secure and distributed way to manage transactions between prosumers. Besides, smart contracts combine protocols with user interfaces to formalize and secure relationships over computer networks Szabo, 1997. In practice, smart contracts offer to facilitate and automatize transactions, as soon as requirements are observed and satisfied.

Distributed ledgers combined with smart contracts may offer the opportunity to easily implement energy management systems incorporating adaptive, data-driven controllers where RL might be revealed to be an effective approach.

Dynamics of the deployment of renewable energy production capacities Long-term planning of renewable energy production capacities is the starting point in efficient energy harnessing from these sources. It is natural to formulate this problem as a sequential decision making problem, and RL is an obvious choice to solve this problem. The problem consideration as a dynamic one over a time span of several years opens up the possibility to compute the dynamics (when and the amount) of the renewable energy production capacities deployment, thus maximizing the energy harnessing from renewable sources over the time span.

The notion of Energy Return on Energy Investment (ERoEI) as characteristics of technologies, can be used to this purposes (Fonteneau and Ernst (2017)). ERoEI is the ratio of the amount of usable energy acquired from a particular energy resource to the amount of energy expended to obtain that energy resource.

[1] see `http://brooklynmicrogrid.com/`
[2] see also `www.blockchaintechnologies.com/`

All of the above, together with the development of advanced communications and measurement infrastructure, turns modern power systems into *cyber-physical systems*. Complexity of these systems makes design of the system controllers difficult and RL is expected to take a greater role in designing (learning) appropriate control laws.

Moreover, some distributed generation which is exploiting renewable energy sources inherently brings uncertainties into the system operation and control. Recent work presented in Gao et al. (2016), and the references therein, suggests a solution in the form of an adaptive and optimal output-feedback problem for continuous-time uncertain systems with nonlinear dynamic uncertainties with guaranteed robustness to these uncertainties.

5.3 Perspectives for RL in restorative, emergency, robust and distributed control

In our view, the problems that would permit easy applications of RL methods are those where the parameters of existing controllers have to be determined (and usually averaged) off-line over a set of pre-defined (based on engineering judgement) scenarios. This is usually the case with emergency controls.

One example is emergency undervoltage load shedding for voltage instability introduced in Otomega et al. (2007). The same holds true for so-called system integrity protection schemes (SIPS) Madani et al. (2010) since they are designed off-line with the purpose to serve as an emergency control.

Restorative control was considered in the past for transmission systems Ye et al., 2011. With the emergence of renewable energy generation the same needs exist for distribution systems and microgrids. Implementation details might vary, but the principle is the same as for the transmission system and RL methods could offer a solution for this problem. The same holds true for emergency controls and SIPS.

We believe the efficient fusion of RL and control theory and practice could offer solutions where existing controllers provide baseline stability guarantees, while RL improves performances (an example: robust controls are designed for worst-case conditions but most of the time operate in non-worst-case situations, so their synergy could potentially lead to an efficient RL based controls).

If designed as a multi-agent system, then several control agents using RL are connected to a single electric power system. These agents are able to learn in parallel and adapt their performances progressively, leading to a coordinated distributed control. We also envision the great potentials of RL in this direction. In this respect we also emphasize a specifics of power systems, i.e. a sort of communication infrastructure already exists since all substations are connected by electrical lines and any action of one agent is sensible by other ones (at least the ones in close electrical distance) which facilitates the design of multi-agent systems.

5.4 Perspectives for RL in electric power system dynamic security assessment and control

Preventive controls in a normal system operating state (see Fig. 1) are computed using so-called security assessment where controls are computed by applying most-probable disturbances (also known as credible disturbances (contingencies)) and simulating a system response to these disturbances. The disturbances usually considered include short circuits followed by the outages of most-impacted generation and lines. These controls are applied in normal mode and might be revealed to be costly since disturbance actually (and often) does not happen.

There are some possibilities to combine preventive and corrective controls (corrective controls take place when a disturbance happens). This was recognized some time ago in Wehenkel et al. (2006); Ruiz-Vega et al. (2003). An automatic learning was suggested as a possible approach in Wehenkel et al. (2006) while Ruiz-Vega et al. (2003) suggest a combination of open-loop and closed-loop techniques for transient stability control. We believe it is worth revisiting this idea and considering preventive and emergency controls as a single problem, or to use RL to deal with preventive and open-loop controls. The problem boils down to computing the trade-off of incurring costs in prevention with expected costs in emergency control Wehenkel et al. (2006); Ruiz-Vega et al. (2003). The optimal combination of these controls is essentially a sequential decision problem in uncertain environments to which RL methods could provide an efficient solution.

6. CONCLUSION

Based on the extensive review of past considerations of RL methods for electric power system control and decision, the search for new opportunities and our experience working in both RL and power systems fields, we draw the following conclusions:

- RL was already considered as an effective approach to solve many electric power system control and decision problems confirming the potential of RL methods and interest of the power system research community in considering these methods.
- RL methods could be extended to other power system problems, most notably restorative and emergency control to be used not only by transmission but also distribution system and microgrid operators, as well as unit commitment (see Dalal and Mannor (2015) and references therein). Emerging devices such as storage, electric vehicles and power electronics devices should also be considered in this context.
- Recent breakthroughs in RL offer solutions that handle data sets more efficiently and open up possibilities for using RL in on-line mode. Most relevant breakthroughs for electric power system control and decision are Safe RL, Deep RL and path integral control for RL. Past considerations are worth revisiting in the context of these new RL methods together with their use for other potentially interesting power system problems.
- There is a need for better fusion of control theory and practice with RL while designing power system

controllers. We mention the robust control and fusion with stability oriented controls as good examples in this context, and path integral control for RL opens up new possibilities in this context.
- Embedding domain-specific knowledge makes many RL for power system control/decision tractable. This requires closer cooperation between those interested in RL methods and power engineers in solving many practical problems. This will certainly contribute to a wider considerations of RL methods for power system problems. We suggest this as the right way to go with practical implementations of the research results in this field. In particular, let us mention Bayesian RL approaches (see Ghavamzadeh et al. (2015) for an extensive literature review), which offer two interesting features: by assuming a prior distribution on potential (unknown) environments, Bayesian RL (i) allows to formalize Bayesian-optimal exploration / exploitation strategies, and (ii) offers the opportunity to incorporate prior knowledge into the prior distribution. However, most Bayesian RL algorithms suffer computational complexity (Castronovo et al. (2016)).
- Potential considerations of RL are indeed not limited to those recommended by the authors (recommendations of this paper are just reflections of the authors working in both fields of RL and electric power systems). In principle, any sequential decision problem, regardless of the entity to use it (transmission or distribution system operator, microgrid operator, retailer, load aggregator, or distribution generation owners), is worthy of RL consideration for its solution.

Modern power systems, with a high proliferation of distributed generation, new load types, increased use of power electrics devices, new measurement technologies and upgrade of communications infrastructure are cyber-physical systems. It is reasonable to expect increased use of RL (and other machine learning techniques) since the complexity of cyber-physical systems makes control and decision difficult without the learning techniques.

In general, we strongly encourage further consideration of RL for the solution of electric power system decision and control problems, either through revisiting already considered ones, in light of the use of new RL methods, or as an extension to new problems.

ACKNOWLEDGEMENTS

The authors thank the Walloon Region who has funded this research in the context of the BATWAL project.

REFERENCES

Ademoye, T. and Feliachi, A. (2012). Reinforcement learning tuned decentralized synergetic control of power systems. *Elec. Power Syst. Research*, 86, 34–40.

Ahamed, T.P.I., Rao, P.S.N., and Sastry, P.S. (2002). A reinforcement learning approach to automatic generation control. *Elec. Power Syst. Research*, 63, 9–26.

Aittahar, S., François-Lavet, V., Lodeweyckx, S., Ernst, D., and Fonteneau, R. (2015). Imitative learning for online planning in microgrids. In *International Workshop on Data Analytics for Renewable Energy Integration*, 1–15. Springer International Publishing.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, 2, 1–127.

Busoniu, L., Babuska, R., Schutter, B.D., and Ernst, D. (2010). *Reinforcement learning and dynamic programming using function approximators.* CRC Press, Boca Raton.

Castronovo, M., Ernst, D., Couëtoux, A., and Fonteneau, R. (2016). Benchmarking for bayesian reinforcement learning. *PloS one*, 11(6), e0157088.

Dalal, G., Gilboa, E., and Mannor, S. (2016). Hierarchical decision making in electricity grid management. In *Proceedings of The 33rd International Conference on Machine Learning*, 2197–2206.

Dalal, G. and Mannor, S. (2015). Reinforcement learning for the unit commitment problem. In *PowerTech, 2015 IEEE Eindhoven*, 1–6. IEEE.

Daneshfar, F. and Bevrani, H. (2010). Loadfrequency control: a GA-based multi-agent reinforcement learning. *IET Gen., Transm., Dist.*, 4, 13–26.

Dubois, A., Wehenkel, A., Fonteneau, R., Olivier, F., and Ernst, D. (2017). An app-based algorithmic approach for harvesting local and renewable energy using electric vehicles. In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)*, 1–6. ICAART.

DyLiacco, T.E. (1974). Real-time computer control of power systems. *Proc. IEEE*, 62, 884–891.

Ekanayake, J., Liyanage, K., Wu, J., and Jenkins, N. (2012). *Smart Grid: Technology and Applications.* Wiley, Chichester.

Ernst, D., Glavic, M., Capitanescu, F., and Wehenkel, L. (2009). Reinforcement learning versus model predictive control: A comparison on a power system problem. *IEEE Trans. Syst., Man, Cyber.: Part B*, 39, 517–529.

Ernst, D., Glavic, M., and Wehenkel, L. (2004). Power systems stability control: Reinforcement learning framework. *IEEE Trans. Power Syst.*, 19, 427–435.

Fonteneau, R. and Ernst, D. (2017). On the dynamics of the deployment of renewable energy production capacities. In *Mathematical Advances Towards Sustainable Environmental Systems*, 23–60. Springer.

Fonteneau, R., Ernst, D., Boigelot, B., and Louveaux, Q. (2013a). Min max generalization for deterministic batch mode reinforcement learning: relaxation schemes. *SIAM Journal on Control and Optimization*, 51(5), 3355–3385.

Fonteneau, R., Murphy, S.A., Wehenkel, L., and Ernst, D. (2013b). Batch mode reinforcement learning based on the synthesis of artificial trajectories. *Annals of Operations Research*, 208(1), 383–416.

François-Lavet, V., Gemine, Q., Ernst, D., and Fonteneau, R. (2016a). Towards the minimization of the levelized energy costs of microgrids using both long-term and short-term storage devices. *Smart Grid: Networking, Data Management, and Business Models*, 295–319.

François-Lavet, V., Taralla, D., Ernst, D., and Fonteneau, R. (2016b). Deep reinforcement learning solutions for energy microgrids management. In *European Workshop on Reinforcement Learning.*

Gao, W., Jiang, Y., Jiang, Z.P., and Chai, T. (2016). Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming. *Automatica*, 72, 37–45.

Garcia, J. and Fernandez, F. (2015). A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16, 1437–1480.

Gemine, Q., Ernst, D., and Cornélusse, B. (2014). Active network management for electrical distribution systems: problem formulation, benchmark, and approximate solution. *arXiv preprint arXiv:1405.2806*.

Ghavamzadeh, M., Mannor, S., Pineau, J., Tamar, A., et al. (2015). Bayesian reinforcement learning: A survey. *Foundations and Trends® in Machine Learning*, 8(5-6), 359–483.

Glavic, M. (2005). Design of a resistive brake controller for power system stability enhancement using reinforcement learning. *IEEE Trans. Contr. Syst. Tech.*, 13, 743–751.

Glavic, M. and Alvarado, F. (2016). Potential, opportunities and benefits of electric vehicles as frequency regulation resources. In *Open Repository and Bibliography (ORBi: https://orbi.ulg.ac.be/), working paper*, 1–10. The University of Liege, Belgium.

Glavic, M., Ernst, D., and Wehenkel, L. (2005a). Combining a stability and a performance-oriented control in power systems. *IEEE Trans. Power Syst.*, 20, 525–526.

Glavic, M., Ernst, D., and Wehenkel, L. (2005b). A reinforcement learning based discrete supplementary control for power system transient stability enhancement. *Engineering Intelligent Systems for Electrical Engineering and Communications*, 13, 81–88.

Glavic, M., Ernst, D., and Wehenkel, L. (2006). Damping control by fusion of reinforcement learning and control Lyapunov functions. In *Proc. of f the 38th North American Power Symposium (NAPS 2006)*, 1–7. NAPS.

Hadidi, R. and Jeyasurya, B. (2013). Reinforcement learning based real-time wide-area stabilizing control agents to enhance power system stability. *IEEE Trans. Smart Grid*, 4, 489–497.

Harp, S.A., Brignone, S., Wollenberg, B.F., and Samad, T. (2000). SEPIA: A simulator for electric power industry agents. *IEEE Contr. Syst. Mag.*, 20, 53–69.

Hasselt, D.V., Guez, A., and Silver, D. (2015). Deep reinforcement learning with double Q-learning. *CoRR, abs/1509.06461*.

Hasselt, H.V. (2010). Double Q-learning. In *Advances in Neural Information Processing Systems*, 2613–2621.

Jasmin, E.A., Ahamed, T.P.I., and Raj, V.P.J. (2011). Reinforcement learning approaches to economic dispatch problem. *Int. Journal Elec. Power and Ener. Syst.*, 33, 836–845.

Karimi, A., Eftekharnejad, S., and Feliachi, A. (2009). Reinforcement learning based backstepping control of power system oscillations. *Elec. Power Syst. Research*, 79, 1511–1520.

Khorramabady, S.S., Bakhshai, A., and Bakhshai, A. (2015). Intelligent control of grid-connected microgrids: An adaptive critic-based approach. *IEEE Trans. Emer. Selec. Topics Power Electr.*, 3, 493–504.

Kim, B.G., Zhang, Y., van der Schaar, M., and Lee, J.W. (2016). Dynamic pricing and energy consumption scheduling with reinforcement learning. *IEEE Trans. Smart Grid*, 7, 2187–2198.

Krause, T., Beck, E.V., Cherkaoui, R., Germond, A., Andersson, G., and Ernst, D. (2006). A comparison of nash equilibria analysis and agent-based modelling for power markets. *Int. Journal Elec. Power and Ener. Syst.*, 28, 599–607.

Kydd, R.H., Anstrom, J.R., Heitmann, P.D., Komara, K.J., and Crouse, M.E. (2016). Vehicle-solar-grid integration: Concept and construction. *IEEE Power Ener. Tech. Syst. Journal*, 3, 81–88.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.

Li, B.H. and Wu, Q.H. (1999). Learning coordinated fuzzy logic control of dynamic quadrature boosters in multimachine power systems. *IEE Proc. Gen., Transm., Dist.*, 6, 577–585.

Li, F.D., Wu, H., He, Y., and Chen, X. (2012). Optimal control in microgrid using multi-agent reinforcement learning. *ISA Transactions*, 51, 743–751.

Lincoln, R., Galloway, S., Stephen, B., and Burt, G. (2012). Comparing policy gradient and value function based reinforcement learning methods in simulated electrical power trade. *IEEE Trans. Power Syst.*, 27, 373–380.

Madani, V., Novosel, D., Horowitz, S., Adamiak, M., Amantegui, J., Karlsson, D., Imai, S., and Apostolov, A. (2010). IEEE PSRC report on global industry experiences with system integrity protection schemes (SIPS). *IEEE Trans. Power Syst.*, 25, 2143–2155.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518, 529–533.

Mohagheghi, S., Venayagamoorthy, G.K., and Harley, R.G. (2006). Adaptive critic design based neuro-fuzzy controller for a static compensator in a multimachine power system. *IEEE Trans. Power Syst.*, 21, 1744–1754.

Nanduri, V. and Das, T.K. (2007). A reinforcement learning model to assess market power under auction-based energy pricing. *IEEE Trans. Power Syst.*, 22, 85–95.

Olivier, F., Aristidou, P., Ernst, D., and Van Cutsem, T. (2016). Active management of low-voltage networks for mitigating overvoltages due to photovoltaic units. *IEEE Transactions on Smart Grid*, 7(2), 926–936.

Otomega, B., Glavic, M., and Van Cutsem, T. (2007). Distributed undervoltage load shedding. *IEEE Trans. Power Syst.*, 22, 2283–2284.

Padiyar, K.R. (ed.) (2008). *Power System Dynamics, Stability and Control*. BS Publications.

Rahimiyan, M., Mashhadi, H.R., and Mashhadi, H.R. (2010). An adaptive Q-learning algorithm developed for agent-based computational modeling of electricity market. *IEEE Trans. Syst., Man, Cyber.: Part C*, 40, 547–556.

Rochet, J.C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1, 990–1029.

Ruelens, F., Claessens, B.J., Vandael, S., Schutter, B.D., Babuska, R., and Belmans, R. (2016). Residential demand response of thermostatically controlled loads using batch reinforcement learning. *IEEE Trans. Smart Grid*, In Press, 1–11.

Ruiz-Vega, D., Glavic, M., and Ernst, D. (2003). Transient stability emergency control combining open-loop and closed-loop techniques. In *IEEE PES General meeting*, 2053–2059. IEEE.

Schaul, T., Quan, J., Antonoglou, I., and Silver, D. (2015). Prioritized experience replay. *arXiv preprint arXiv:1511.05952*.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Driessche, G.V.D., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, 484–489.

Sutton, R.S. and Barto, A.G. (1998). *Reinforcement Learning: An Introduction.* MIT Press, Boston.

Szabo, N. (1997). Formalizing and securing relationships on public networks. *Peer-Reviewed Journal on Internet*, 2.

Tang, Y., He, H., Wen, J., and Liu, J. (2015). Power system stability control for a wind farm based on adaptive dynamic programming. *IEEE Trans. Smart Grid*, 6, 166–177.

Theodorou, E.A., Buchli, J., and Schaal, S. (2010). A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, 11, 3137–3181.

Thomas, P.S. (2015). *Safe Reinforcement Learning.* PhD Dissertation, The University of Massachusetts Amherst, Amherst.

US-DoE (2004). Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations. Technical report, US Department of Energy, US-Canada Power System Outage Task Force.

Vandael, S., Claessens, B., Ernst, D., Holvoet, T., and Deconinck, G. (2015). Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market. *IEEE Trans. Smart Grid*, 6, 1795–1805.

Venayagamorthy, G.K., Sharma, R.K., Gautam, P.K., and Ahmadi, A. (2016). Dynamic energy management system for a smart microgrid. *IEEE Trans. Neural Networks and Learning Systems*, 27, 1643–1656.

Vlachogiannis, J.G., Hatziargyriou, N., and Hatziargyriou, N. (2004). Reinforcement learning for reactive power control. *IEEE Trans. Power Syst.*, 19, 1317–1325.

Wang, D., Glavic, M., and Wehenkel, L. (2014). Trajectory-based supplementary damping control for power system electromechanical oscillations. *IEEE Trans. Power Syst.*, 29, 2835–2845.

Wang, Y., Lin, X., and Pedram, M. (2016). A near-optimal model-based control algorithm for households equipped with residential photovoltaic power generation and energy storage systems. *IEEE Trans. Sust. Ener.*, 7, 77–86.

Watkins, C.J.C.H. and Dayan, P. (1992). Q-learning. *Machine learning*, 8, 279–292.

Wehenkel, L., Glavic, M., Geurts, P., and Ernst, D. (2006). Automatic learning of sequential decision strategies for dynamic security assessment and control. In *IEEE PES General meeting*, 1–6. IEEE.

Wei, C., Zhang, Z., Qiao, W., and Qu, L. (2015). Reinforcement-learning-based intelligent maximum power point tracking control for wind energy conversion systems. *IEEE Trans. Ind. Ecetr.*, 62, 6360–6370.

Xu, Y., Zhang, W., Liu, W., and Ferrese, F. (2012). Multiagent-based reinforcement learning for optimal reactive power dispatch. *IEEE Trans. Syst., Man, Cyber.: Part C*, 42, 1742–1751.

Yan, J., He, H., Zhong, X., and Tang, Y. (2016). Q-learning based vulnerability analysis of smart grid against sequential topology attacks. *IEEE Trans. Inf Forens. Secur.*, In Press, 1–11.

Ye, D., Zhang, M., and Sutanto, D. (2011). A hybrid multiagent framework with Q-learning for power grid systems restoration. *IEEE Trans. Power Syst.*, 26, 2434–2441.

Yousefian, R., Kamalsadan, S., and Kamalsadan, S. (2016). Design and real-time implementation of optimal power system wide-area system-centric controller based on temporal difference learning. *IEEE Trans. Ind. App.*, 52, 395–406.

Yu, T., Wang, H.Z., Zhou, B., Chen, K.W., and Tang, J. (2012a). Multi-agent correlated equilibrium $Q(\lambda)$ learning for coordinated smart generation control of interconnected power grids. *IEEE Trans. Power Syst.*, 27, 373–380.

Yu, T., Zhang, X.S., Zhou, B., and Chan, K.W. (2016). Hierarchical correlated Q-learning for multi-layer optimal generation command dispatch. *Int. Journal Elec. Power and Ener. Syst.*, 78, 1–12.

Yu, T., Zhou, B., Chan, K.W., Chen, L., and Yang, B. (2011). Stochastic optimal relaxed automatic generation control in non-markov environment based on multi-step $Q(\lambda)$ learning. *IEEE Trans. Power Syst.*, 26, 1272–1282.

Yu, T., Zhou, B., Chan, K.W., Yuan, Y., Yang, B., and Wu, Q. (2012b). $R(\lambda)$ imitation learning for automatic generation control of interconnected power grids. *Automatica*, 48, 2130–2136.

Zarabbian, S., Belkacemi, R., and Babalola, A.A. (2016). Reinforcement learning approach for congestion management and cascading failure prevention with experimental application. *Elec. Power Syst. Research*, 141, 179–190.