Mise à jour de l'état de l'art sur les méthodes et outils innovants pour le traitement des systèmes complexes et benchmarking

Julien Niol, Mouna Rifi, Jean Caire, Carole Duval, Albin Tarrisse, Mohamed Hibti, Florent Brissaud

Institut pour la Maîtrise des Risques
Sûreté de Fonctionnement - Management - Cindyniques

Mise à jour de l'état de l'art sur les méthodes et outils innovants pour le traitement des systèmes complexes et *benchmarking*

NIOL Julien
Airbus Protect
1, boulevard Jean Moulin
78990 Elancourt
julien.niol@apsys-airbus.com

DUVAL Carole
EDF R&D
7, boulevard Gaspard Monge
91120 Palaiseau
carole.duval@edf.fr

HIBTI Mohamed
EDF R&D
7, boulevard Gaspard Monge
91120 Palaiseau
mohamed.hibti@edf.fr

RIFI Mouna
EDF R&D
7, boulevard Gaspard Monge
91120 Palaiseau
mouna.rifi@edf.fr

TARRISSE Albin
INERIS
Parc technologique Alata
60550 Verneuil-en-Halatte
albin.tarrisse@ineris.fr

BRISSAUD Florent
RICE GRTgaz
1 Rue du Commandant d'Estienne
d'Orves, 92390 Villeneuve-la-Garenne
florent.brissaud@grtgaz.com

CAIRE Jean
RATP
Paris, France
jean.caire@ratp.fr

Résumé — L'objectif du projet IMdR P20-1 est d'intégrer de nouvelles méthodes pour traiter les systèmes complexes et faire un *benchmark* sur un cas d'usage. Le projet s'est intéressé aux jumeaux numériques selon une démarche MBSE/MBSA, à la théorie des réseaux complexes et aux modèles de systèmes vivants.

Le cas d'usage retenu porte sur une unité de production d'hydrogène. L'installation a fait l'objet d'un modèle MBSA en AltaRica avec l'outil SimfiaNeo. Ce modèle se veut unificateur en permettant l'étude des objectifs de production et de sûreté de fonctionnement, en intégrant les vulnérabilités de cybersécurité. Ce modèle a ensuite été transposé en réseaux pour en étudier des indicateurs de centralité.

Ceci a permis de calculer des indicateurs classiques mais surtout de s'interroger sur les apports d'autres disciplines, avec la nécessité d'approfondir celles-ci, en l'état actuel des connaissances.

Mots-clefs — systèmes complexes, jumeaux numériques, théorie des réseaux, modélisation du vivant, réseaux de neurones

Abstract— The goal of the IMdR P20-1 project is to integrate new methods managing complex systems and to make a benchmark on a use case. The project interested in digital twins under MBSE/MBSA approaches, in complex networks theory and living systems modeling.

The use case selected deals with a hydrogen production unit. The unit has been modeled with AltaRica and SimfiaNeo tool in MBSA. This model ambitions to unify analyses of production and safety goals integrating cybersecurity vulnerabilities. This model has then been transposed in networks to study centrality indicators.

The model allowed computing classical indicators but mainly to discuss on other disciplines contributions with the necessity to go deeper on it, according to current state of the art.

Keywords — complex systems, digital twin, network theory, living systems modeling, neural networks

I. CONTEXTE

Le projet P11-4 de l'IMdR [1], réalisé par APSYS en 2012, dressait déjà un état de l'art sur les différentes méthodes et outils pour mener à bien des analyses de sûreté de fonctionnement de systèmes complexes [2], [3].

Aujourd'hui, sur l'impulsion du groupe de travail et de réflexion (GTR) « Innovation en Rupture Transversale – Modélisation des Systèmes Complexes », l'IMdR fait réaliser par APSYS, après consultation, le projet P20-1 consistant en une mise à jour de l'état de l'art du projet P11-4 afin d'y intégrer des nouvelles méthodes/outils, plus particulièrement les suivants :

- les méthodes de « jumeaux numériques » largement adressées par les démarches *Model-Based System Engineering* et *Model-Based Safety Assessment* (MBSE/MBSA) ;
- la théorie des réseaux développée notamment avec le Laboratoire d'Informatique de Paris-Nord (LIPN) et implémentée pour une meilleure exploitation des Etudes Probabilistes de Sûreté [4], [5] ;
- le « *deep learning* » sur des réseaux de neurones alimentés par du big data présenté par la société DCBrain [6];
- les modélisations du vivant [7], du projet BRAINS et de l'immunologiste Véronique Thomas-Vaslin [8].

Le projet se fixe l'objectif d'identifier les apports de ces méthodes et leurs éventuelles insuffisances ou inconvénients

pour définir des pistes de développement de méthodes innovantes aptes au traitement de systèmes complexes.

Après une première phase d'état de l'art, il a été décidé d'appliquer certaines d'entre elles à un cas d'usage retenu par les souscripteurs du projet, portant sur l'étude d'une installation de production et de stockage d'hydrogène.

Le projet développe dans ce cadre trois sujets qui représentent actuellement des lignes de force majeures dans le champ de l'innovation :

- Les techniques et outils issus des sciences du vivant permettant d'adresser des facteurs de complexité moins classiquement maîtrisés dans le domaine industriel comme les phénomènes d'émergence positive ou négative, les mécanismes de résilience ou encore les capacités régénératives et de mutation rendues possibles par les propriétés intrinsèques des organismes vivants et leurs échanges avec les écosystèmes où ils se développent
- Les approches « MBSE/MBSA » en associant différents modèles (en particulier un modèle d'architecture d'une installation ou d'un système) et abstractions complémentaires dans la direction du « jumeau numérique » (maquette virtuelle d'un système simulant toutes les phases de son cycle de vie), permettent de maîtriser différents points de vue de performances caractérisant un système, en virtualisant les modes de comportements et interactions avec l'environnement extérieur : il s'agit de la continuité de service ou disponibilité opérationnelle, de la fiabilité intrinsèque, de la sécurité ou cybersécurité mais aussi des performances environnementales voire de durabilité qui ne cessent d'étendre le tableau de bord des métriques.
- Les approches fondées sur la théorie des réseaux complexes permettant d'adresser des facteurs d'importance au sein de celui-ci par le calcul des indicateurs de centralité (formulation mathématique de différentes notions de « poids » caractérisant l'importance topologique d'un nœud au sein d'un réseau), dans le but d'identifier les éléments les plus influents sur les propriétés du système y compris ceux inclus dans des évènements rares.

Ces méthodes permettent à la fois de prévoir la sécurité, disponibilité et performance des systèmes au sein des environnements où ils sont déployés, mais également d'optimiser les choix de conception, de développement, d'exploitation ou de maintenance, en fonction des interactions qu'ils subissent et de l'évolution des environnements où ils sont plongés.

II. Methodes, Techniques et Outils issus des Sciences du Vivant

La démarche de recherche de méthodes innovantes qui constituait la spécificité de ce projet, a été conduite dans de nombreux domaines, en particulier à travers le prisme des sciences du vivant. Le projet a essayé d'évaluer si les méthodes et outils utilisés dans ce domaine pouvaient permettre de mettre en lumière de nouvelles propriétés sur les systèmes complexes rencontrés dans l'industrie. Ce travail a néanmoins fait émerger une question essentielle, la définition d'un modèle, avant même de sélectionner des outils susceptibles d'être « expérimentés » à travers un cas d'usage.

Dans le domaine du vivant, il est possible de s'interroger sur les modélisations envisageables sur le plan théorique. Comment chacun va-t-il se représenter la vision du vivant dès lors qu'une fonction a été identifiée, une fonction que l'on veut comprendre voire simuler ? Ensuite, il pourra se présenter la problématique de modèles plus ou moins partageables avec d'autres.

La nature n'offre pas de lois générales. Au contraire de la sociologie, avec ses règles et ses contraintes, les systèmes vivants créent tellement de diversité, qu'ils peuvent échapper à ces lois et contraintes. Les systèmes se dupliquent, se répliquent en multipliant les variantes, s'effondrent, se réorganisent, sans parler des systèmes chaotiques avec des bifurcations liées à l'instabilité.

Un exemple significatif de modèles utilisés dans les sciences du vivant inclut les modèles empiriques et expérimentaux, sur des animaux ou des matériels vivants en général. Au niveau pratique, il s'agit de modélisations et de simulations informatiques de différentes natures qui intègrent les résultats obtenus à partir d'expériences sur des organismes vivants dans leur intégralité qualifiées d'*in vivo*, ainsi que sur des expériences ciblées hors d'un organisme vivant qualifiées d'*in vitro*. Il y a différentes typologies de modèles possibles, selon les objectifs et les questions que l'on se pose et les niveaux de décomposition.

A quoi le terme modèle se réfère-t-il dans les sciences du vivant ? Cette question fait l'objet d'ouvrages entiers dans le domaine du vivant. L'appellation "modèles" recouvre de nombreuses notions, parmi lesquelles des modèles théoriques, des modèles expérimentaux basés sur des modèles vivants ou *in vitro*, permettant de quantifier par des approches analytiques certaines entités. Ils renvoient pour la plupart à des approches réductionnistes rassemblant toutes les approches analytiques basées sur la décomposition des organismes sous forme d'ontologies et les décomposant en systèmes minimalistes que l'on va pouvoir analyser, dénombrer et en étudier les interactions. L'analyse passe toujours par la décomposition sous forme de petits modules à travers des arborescences. Il s'agit de passer des phénotypes des fonctions aux cellules. C'est une démarche descriptive, qualitative ou quantitative qui peut aboutir par exemple au séquençage massif des génomes. Malgré tout, il apparait rapidement que ce n'est pas en « coupant le vivant en petit morceaux » que cela apportera des réponses suffisantes pour appréhender sa complexité et sa diversité.

En tout état de cause, la robustesse, la résistance et la résilience sont trois facteurs stratégiques pour expliquer la survie des espèces. La diversité joue ici également un rôle clef, cela donne de la créativité. Le système global sera peut-être moins efficace, mais plus résistant. Cependant, il faut de l'historique pour que cela « marche » car la transmission de la mémoire est clef.

Il faut citer également les modèles en réseaux dynamiques qui fonctionnent dans l'espace et le temps avec différentes couches, intégrant des milliers de sous-systèmes entremêlés. Il est nécessaire de décrire l'entremêlement élémentaire. Par exemple, une des lois concerne les gènes qui produisent un certain déterminisme. Mais ce déterminisme n'est pas absolu. Ainsi, deux vrais jumeaux ont en effet plus de 80% de différences sur leurs lymphocytes., qui sont les cellules

chargées de mettre en œuvre les fonctions d'immunité et de défense vis-à-vis des agressions externes telles que les virus, bactéries ou internes dues à des dérèglements ou anomalies liées à la prolifération de cellules de type cancéreux…

La plupart des modèles traite des objets élémentaires et de réseaux en renouvellement permanent. Il y a en fait deux niveaux de dynamiques, une dynamique entre les niveaux et une dynamique de renouvellement à chaque niveau, à laquelle il faut rajouter une dynamique de transmission. Par exemple, une cellule mère donne deux cellules filles qui sont différentes.

Cependant, concernant les réseaux dynamiques, le modèle *in vivo* est le plus éloquent : rappelons qu'un modèle *in vivo* renvoie à une expérimentation effectuée sur un système vivant dans son ensemble, alors qu'un modèle *in vitro* renvoie à une expérimentation effectuée sur une partie du système vivant, toutes choses étant reconstituées « à l'identique » par ailleurs, dans la mesure du possible…

Ainsi, à travers les réseaux dynamiques supportant la cognition des cellules, les systèmes peuvent s'exprimer dans leur vraie grandeur *in vivo*. *In vitro*, la complexité est atténuée. Par exemple, la cognition des cellules n'est pas aussi forte *in vitro*. Il n'y a pas de mémoire dans une boîte de culture. Les réseaux sont simplifiés et perdent de leur dynamique en extrayant les cellules de l'organisme.

Vu l'importance des réseaux dans le monde du vivant, et de leur rôle prépondérant dans la transmission des informations et le conditionnement et le comportement des populations cellulaires, il nous a donc paru intéressant d'illustrer par un cas d'usage concret la puissance de conceptualisation des modèles de type réseau et leur richesse de caractérisation et de traitement du point de vue par exemple des facteurs de complexité qu'ils permettent d'illustrer et d'évaluer.

Le paragraphe suivant présente ce cas d'usage, exploité dans la suite de l'article, à travers le champ de deux méthodes innovantes d'analyse système :

- L'application d'une approche « MBSE/MBSA » (Model Based System Engineering/Model Based Safety Analysis) « unifiée » dont les résultats préfigurent ce que pourront être les analyses unifiées de type « jumeau numérique » basées sur des maquettes virtuelles adressant d'emblée toutes les métriques pertinentes du type sécurité, cyber sécurité, résilience, et disponibilité opérationnelle de production par exemple
- L'application d'une approche purement « réseaux » permettant de mettre en relation avec les métriques précédentes, plus classiques, toute la palette des facteurs de centralité et autres indices de complexité offerte par le référentiel mathématique de la théorie des réseaux…

III. CAS D'USAGE

A. *Périmètre du cas d'usage*

Le cas d'usage porte sur une unité de production et de stockage d'hydrogène vert, dont le schéma est présenté en Fig. 1. Cette unité a été simplifiée pour ne conserver que les grandes étapes de production sans détailler tous les composants présents, dans le but d'expérimenter sur un modèle de taille raisonnable avant d'envisager un passage à une plus grande échelle.

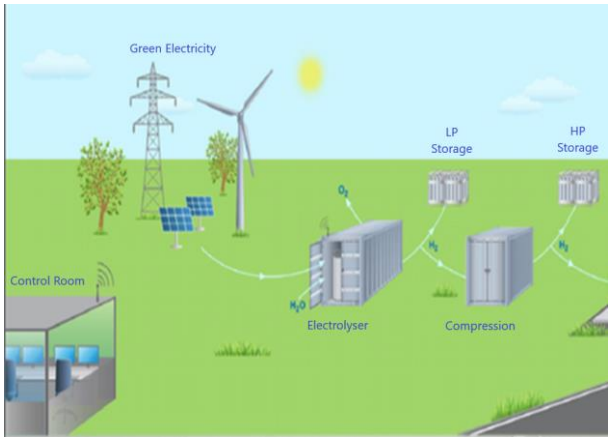

Fig. 1. Schéma de principe de l'installation (source : INERIS)

La production d'hydrogène est assurée par électrolyse de l'eau. L'hydrogène ainsi produit est stocké dans plusieurs réservoirs avec des pressions de stockage différentes avant d'être exporté vers les consommateurs finaux. Le modèle se compose de l'électrolyseur et de deux réservoirs de stockage, ainsi que des pompes et compresseurs permettant d'assurer le transfert d'un composant à l'autre.

L'hydrogène est qualifié de vert car l'électricité employée lors de l'électrolyse est produite à partir d'une source d'énergie renouvelable. Le cas d'usage inclut les dispositifs de production électrique que sont les éoliennes et les panneaux solaires, ainsi que la batterie permettant de gérer les variations de production. L'ensemble de l'installation est relié à un centre de contrôle qui permet le pilotage de l'électrolyse en tenant compte à la fois, de l'énergie disponible, de la demande des consommateurs et des stocks d'hydrogène.

L'installation est également représentée dans son environnement immédiat, ayant à son voisinage d'autres sites industriels ou des zones résidentielles qui peuvent être affectées par une situation accidentelle, par exemple l'explosion d'un réservoir d'hydrogène. Le cas d'usage ne se limite pas à représenter la seule installation mais inclut également les conséquences de ces situations sur les cibles voisines. De manière identique, certaines situations accidentelles d'une installation voisine, par effet, peuvent affecter l'unité de production d'hydrogène.

D'autre part, l'installation est pilotée à distance depuis une salle de contrôle centralisée. Il se produit des échanges de données et de commandes entre les différents équipements et cette salle. L'exploitation s'appuie sur les données mesurées par les capteurs pour piloter les différents actionneurs à leur disposition. Ces flux de données constituent des points d'entrée pour la cyber malveillance ce qui est également pris en compte dans le cas d'usage.

B. *Enjeux de l'étude*

Une installation industrielle présente plusieurs enjeux de sécurité et de disponibilité dont il faut tenir compte dès la conception de l'installation. Les méthodes et outils étudiés dans le cadre de l'analyse des systèmes complexes ont pour but d'aider à cette prise en compte.

Un premier enjeu considéré est celui de la disponibilité de production. La fonction principale de l'installation est la production d'hydrogène. Celle-ci se doit donc de répondre au

à la demande des consommateurs en fournissant ce qui est demandé. Le dimensionnement de l'installation ainsi que la stratégie de pilotage et de maintenance sont importants pour atteindre une disponibilité de production la plus élevée possible tout en limitant le coût total de l'installation. Les outils de modélisation peuvent ainsi aider à rechercher un optimum entre ces deux objectifs.

Un deuxième enjeu considéré est celui de la sécurité de l'installation. Celle-ci fait partie des ICPE (Installations Classées Protection de l'Environnement) qui sont soumises à déclaration ou à autorisation auprès de l'administration et qui doivent satisfaire plusieurs exigences réglementaires pour être autorisées à exploiter. Au sein de celles-ci, il est nécessaire d'évaluer les risques pouvant impacter le voisinage du site et d'en démontrer la maîtrise pour limiter le risque de manière acceptable. Les outils de modélisation peuvent aider à évaluer les conséquences des phénomènes dangereux ainsi qu'à rechercher les séquences menant à l'accident pour inclure des barrières adaptées.

Un troisième enjeu considéré est la prise en compte de la vulnérabilité de l'installation à la cyber malveillance. Selon la nature de l'attaque, celle-ci peut mener à diverses conséquences, allant de la perte de production, affectant le premier enjeu, au déclenchement d'une séquence accidentelle affectant le deuxième enjeu. Les outils de modélisation peuvent ainsi aider à appréhender les effets d'une attaque cyber et d'identifier les points de vulnérabilité pour renforcer le système.

Ces enjeux sont déjà largement pris en compte dans l'industrie, chacun disposant de sa boîte à outils spécifique. Au-delà d'une simple étude séparée de chacun de ces aspects, l'intérêt du cas d'usage sélectionné est de pouvoir étudier les méthodes et outils dans une approche commune qui permettrait d'intégrer les dépendances entre chaque discipline et d'étudier le système dans sa globalité.

IV. Modélisation

Pour la modélisation du cas d'usage présenté dans la section précédente, il a été retenu de s'appuyer sur un modèle MBSE/MBSA pour formaliser les éléments contenus dans celui-ci et calculer une première série d'indicateurs.

Des approches complémentaires ont ensuite été étudiées à partir de ce modèle initial, en s'appuyant sur la théorie des réseaux complexes.

A. Modèle unificateur MBSE/MBSA

Le modèle est construit à partir du langage AltaRica et implémenté dans l'outil SimfiaNeo développé par Apsys [9]. AltaRica est un langage formel basé sur des automates de mode que l'on peut assimiler à un système à transitions gardées [10]. Ce choix est une solution parmi d'autres outils et formalismes existants. Le cas aurait aussi pu être traité par l'intermédiaire de l'outil KB3 d'EDF et le langage Figaro ou par l'utilisation d'outils implémentant des réseaux de Petri.

Le modèle représente le cas d'usage et ses principaux équipements sous forme d'une architecture ressemblant à celle qui est dessinée sur un *Process Flow Diagram* (PFD) pour ce genre d'installation, dont le niveau supérieur est visible en Fig. 2. La première couche du modèle est fonctionnelle. Elle capture la logique de fonctionnement des équipements et l'enchainement des étapes de production lorsque tout est nominal. Le modèle est ensuite enrichi d'états

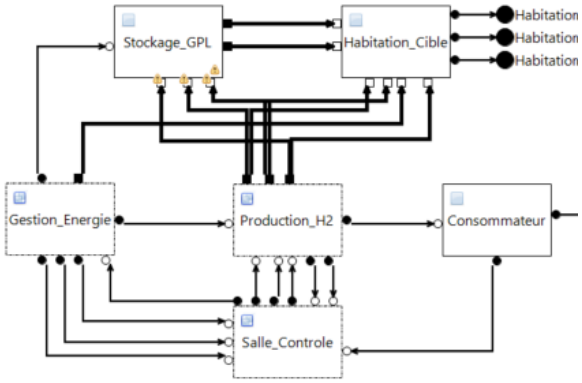

Fig. 2. Niveau supérieur de l'architecture du modèle

défaillants et de leurs évènements initiateurs pour ajouter les transitions menant d'un fonctionnement nominal à un fonctionnement dégradé. Ces évènements initiateurs couvrent à la fois des défaillances de composants, mais aussi des attaques cyber ou des modifications d'environnement qui affecteraient le système. Un exemple de code est donné en Fig. 3, permettant de décrire les effets d'une attaque cyber leurrant le capteur de niveau à une valeur plus basse que la réalité.

La nature du langage AltaRica permet d'exploiter le modèle par l'intermédiaire d'une simulation évènementielle. Lorsqu'un évènement est tiré, les transitions associées sont franchies et l'état global du modèle est mis à jour. L'enchainement des tirages permet de construire des séquences d'évènements depuis la situation nominale et d'évaluer les conséquences de cette séquence sur le système.

L'outil SimfiaNeo propose plusieurs algorithmes pour exploiter le modèle et calculer les indicateurs utiles aux enjeux présentés dans la section précédente.

- La simulation stochastique (Monte-Carlo) : chaque évènement est déclaré avec sa loi de probabilité associée qui permet de construire des histoires probabilisées. Il est ainsi construit un indicateur de disponibilité de production en observant dans le modèle si la quantité produite est suffisante pour répondre aux attentes du consommateur. Le résultat

```
node Salle_Controle__Attaque_Capteur
 flow
        Mesure_HP : Stockage : in;
        Mesure_a_HP : Stockage : out;
 state
        integrite_HP : MDD_Capteur_Niveau ;
 init
        integrite_HP := nominal;
 event
        leurrage_Bas_HP;
 trans
        integrite_HP=nominal |- leurrage_Bas_HP ->
integrite_HP:=sousestime;
 assert
        Mesure_a_HP = case{
        (integrite_HP=sousestime):Bas,
        else Mesure_HP};
edon
```

Fig. 3. Exemple de code AltaRica

est ainsi une valeur de disponibilité de production moyennée sur l'ensemble des histoires.

- La génération de séquences : l'outil explore de manière exhaustive l'espace d'état du modèle pour déterminer l'ensemble des séquences qui mène de la situation initiale à une situation redoutée représentant un sous-ensemble de l'espace d'état. Cette exploration se fait en déclenchant de manière successive les différentes transitions tirables du modèle dans l'état courant et en évaluant à chaque étape si l'état atteint appartient au sous-espace représentant la situation redoutée. Une fois la séquence complétée, l'outil reprend la séquence suivante à partir du précédent point de divergence pour évaluer les transitions qui n'ont pas encore été explorées. Il est ainsi construit un ensemble de séquences accidentelles menant à une situation redoutée précise avec la possibilité de quantifier la probabilité d'occurrence de cette situation redoutée.

- Le calcul de facteurs d'importance : Il est également possible de calculer les principaux facteurs d'importance tels que définis par [11], qui sont le facteur d'importance marginale (ou Birnbaum), le facteur d'importance critique, le Facteur d'augmentation du risque (ou RAW, *Risk Achievement Worth*), le facteur de diminution du risque (ou RRW, *Reliability Reduction Worth*) ainsi que le facteur d'importance de Vesely-Fussell. Ces facteurs d'importance complètent l'analyse globale du système en permettant l'identification des composants jouant un rôle plus central dans le bon fonctionnement du système et sur lesquels il est pertinent de faire porter l'effort d'amélioration ou de fiabilisation.

Les séquences accidentelles sont générées à partir des évènements déclarés dans le modèle. Ces évènements portent sur les différentes disciplines qui sont unifiées au sein du modèle. Ainsi, les séquences accidentelles qui sont générées peuvent contenir à la fois des défaillances matérielles aléatoires mais aussi des attaques cyber sur les données mesurées. Trois exemples pertinents ont pu être identifiés au travers de cette génération.

- Une exploitation de l'électrolyseur alcalin dans un fonctionnement dégradé menant à une déformation de la membrane jusqu'à sa perte d'étanchéité, entrainant la formation d'un mélange oxygène-hydrogène susceptible de s'accumuler dans un stockage de grande contenance jusqu'à produire une explosion aux effets significatifs

- Un leurrage par attaque cyber du capteur de niveau de la cuve de stockage, faisant réagir l'installation par une surproduction d'hydrogène. Dans le cadre d'une installation nécessitant des stockages à différentes pressions selon les applications et munie d'un seul compresseur dimensionné à la pression maximale de stockage, il est possible de surremplir jusqu'à l'éclatement un réservoir dont la pression de service est inférieure à la pression de dimensionnement du compresseur. Cette séquence est amplifiée en cas de défaillance sur la soupape de sécurité qui doit limiter la surpression.

- L'opération dans des conditions environnementales adverses, une tempête, susceptible d'entrainer une rupture de pale d'une éolienne si la survitesse n'est pas correctement détectée et l'éolienne mise en sécurité. La pale est susceptible de retomber dans une zone dont le rayon contient des équipements sous pression ce qui peut entrainer une perte de confinement et une explosion en champ libre de l'hydrogène ainsi rejeté.

Les indicateurs obtenus avec le modèle sont des indicateurs classiques pour la sûreté de fonctionnement, tels que la disponibilité ou la probabilité d'occurrence d'une situation redoutée. La génération de séquences n'est cependant jamais exhaustive en pratique. L'algorithme est conçu pour permettre l'exploration exhaustive de l'espace d'état du modèle, mais celui-ci est souvent trop étendu pour être parcouru dans un temps raisonnable. Il est donc fait l'hypothèse de tronquer les séquences obtenues en limitant la taille de celles-ci à un ordre donné, représentant le nombre d'évènements considérés dans la séquence. Cette troncature repose sur l'hypothèse que les séquences d'ordre important ont une faible probabilité de survenir, négligeable face à la couverture des séquences déjà générées. Ceci pousse à explorer d'autres solutions, en particulier l'analyse topologique de réseaux.

B. Approches complémentaires : analyse topologique de réseaux

La génération des séquences accidentelles menant à un évènement redouté fournit des résultats similaires à une étude probabiliste de sûreté (EPS) réalisée sur une installation nucléaire. Dans le cadre de travaux de thèse [4], ces séquences ont été modélisées en graphes. A partir d'un évènement initiateur, le réseau est construit en modélisant les fonctions de sûreté et les transitions représentant leurs succès ou l'échec prévues dans l'Analyse Qualitative des Séquences accidentelles menant à des conséquences acceptables ou non. On reconstruit ainsi les séquences menant à l'évènement redouté ainsi que toutes les parades mises en place par les fonctions de sûreté permettant d'éviter l'accident ou d'en atténuer les conséquences.

Une analyse topologique de ce réseau est menée en calculant des indicateurs de centralité pouvant faire émerger de nouveaux facteurs d'importance. Les indicateurs de centralité [12] permettent d'évaluer l'importance d'un sommet ou d'un arc dans le réseau en s'appuyant sur les liens entre ces éléments ou des calculs de distance entre eux, qui permettent d'en identifier les plus centraux ou ceux qui apparaissent le plus souvent sur les chemins les plus courts.

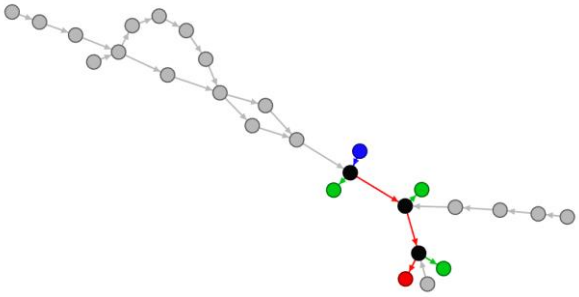

Fig. 4. Visualisation du réseau obtenu

Ces éléments peuvent alors être considérés comme des éléments potentiellement importants pour le système par leur position privilégiée au sein du réseau.

Les réseaux ont été construits à partir des séquences accidentelles générées par le modèle. Pour chaque évènement initiateur, il est possible de retrouver les chemins menant à l'évènement redouté à travers les différentes transitions correspondant à un échec des fonctions de sécurité existantes. Le réseau est ensuite complété en identifiant les composants contribuant à une fonction de sécurité en les rattachant par un arc au sommet représentant celle-ci. Les composants sont modélisés par des sommets et reliés par des arcs permettant de reproduire l'architecture des flux entre les composants qui peuvent être de nature physique ou logique. Dans un premier temps, le réseau a été reconstruit avec une approche binaire, considérant qu'une fonction de sécurité partiellement remplie correspond à un échec, sans introduire de notion d'échec partiel. Ceci permet de s'appuyer sur des réseaux simples au lieu de faire appel à des réseaux multicouches permettant d'explorer par exemple les variantes d'échecs. Il s'agit d'une hypothèse conservatrice qui surestime les effets d'une défaillance d'un composant, recevable dans une analyse de sûreté de fonctionnement. Dans le cadre d'une étude de disponibilité de production, il serait intéressant de distinguer les conséquences sur la productivité selon qu'il s'agisse d'un arrêt total de la ligne ou d'une réduction de sa capacité.

Un réseau, présenté en Fig. 4, a été construit à partir d'une séquence accidentelle générée par le modèle, qui est la première, portant sur la déformation de la membrane et l'explosion du mélange oxygène-hydrogène ainsi formé. Pour cette séquence donnée, le modèle a été affiné afin de prendre en compte tous les composants participant aux boucles de régulation et de sécurité de l'électrolyseur, par rapport à la version initiale qui n'entre pas dans le détail des équipements. L'évènement initiateur apparait en bleu, les fonctions de sécurité en noir, les situations acceptables en vert, les situations redoutées en rouge et les composants en gris. Les liens rouges représentent l'échec des fonctions de sécurité présentes sur la séquence accidentelle et les liens verts leur succès. Sur ce réseau, plusieurs indicateurs de centralité ont été évalués :

- Degré : Le degré correspond au nombre d'arcs entrants ou sortant d'un sommet.
- Proximité (Closeness) : La proximité se base sur le nombre de plus courts chemins qui sortent ou mènent à ce sommet.
- Centralité d'intermédiarité (Betweenness) : L'intermédiarité se base sur le nombre de plus courts chemins qui passent par ce sommet.
- Page Rank : Cette mesure s'appuie sur les algorithmes d'indexation de pages web qui permettent d'identifier les pages les plus pertinentes en fonction des liens qui les référencent.
- Hub et Autorité : Cette mesure identifie les hubs qui sont des sommets avec de nombreux liens sortants (qui référencent de nombreux sommets) et les autorités qui sont, par symétrie, des sommets avec de nombreux liens entrant (qui sont référencés par de nombreux sommets).

Contrairement à l'algorithme de génération de séquences utilisé pour l'exploitation du modèle MBSE-MBSA, il n'est pas nécessaire de tronquer le réseau pour obtenir un résultat dans un temps acceptable. L'algorithme de parcours de l'espace d'état est de complexité exponentielle, ce qui le rend rapidement vulnérable à l'explosion combinatoire. A l'inverse, le calcul de centralité est majoritairement de complexité linéaire.

Le graphe présente également l'avantage de permettre une représentation du système la plus exhaustive possible avec un effort minimal. Les outils classiques de la sûreté de fonctionnement comme le diagramme de fiabilité ou l'arbre de défaillance exigent une analyse détaillée du système pour leur construction. Le résultat est bien souvent tronqué soit par hypothèse que les combinaisons les plus rares sont négligeables, soit par défaut d'identification de certaines combinaisons. Le réseau, une fois la séquence principale identifiée est construit en reproduisant l'architecture physique et logique du système dans une transposition fidèle. Cet avantage est par ailleurs commun au modèle MBSE-MBSA dont la démarche de construction suit la même philosophie.

V. RÉSULTATS

L'exploitation du modèle décrit précédemment s'est faite dans un premier temps dans l'outil SimfiaNeo en utilisant les algorithmes de calcul stochastique et de génération de séquences.

Le calcul stochastique a pour but d'estimer une disponibilité de production intrinsèque en s'appuyant sur les taux de défaillance des équipements ainsi que sur l'impact de ces défaillances sur la capacité de production. Le flux principal est ainsi décrit numériquement et représente la capacité de production instantanée. Cette capacité de production est comparée à la demande formulée pour l'export de la production et moyenné sur l'ensemble des histoires pour calculer l'indicateur. Le résultat brut n'a que peu d'intérêt, car le modèle a été simplifié dans un but de démonstration sur un échantillon représentatif. L'ensemble des modes de défaillance n'a pas été intégré et l'impact de l'instrumentation de contrôle est sous-estimé. De plus, s'agissant d'un modèle évènementiel discret et non d'un modèle hybride, les phénomènes physiques tels que l'évolution du niveau d'un réservoir suivent une approche probabiliste et ne reposent pas sur la résolution des équations différentielles. Il y a donc une perte de représentativité exigeant d'utiliser un autre formalisme pour plus de finesse.

La génération de séquences a permis de rechercher l'ensemble des séquences accidentelles menant à un évènement redouté spécifique. L'évènement redouté retenu correspond à une situation où la cible située dans le voisinage de l'installation se trouve à l'intérieur de la zone d'effets pour un seuil donné. La cible sélectionnée peut être impactée directement par un accident sur l'installation d'hydrogène, mais aussi par effet domino sur le stockage voisin. Une première génération a été effectuée jusqu'à l'ordre 3, soit une séquence maximale de trois évènements successifs pour atteindre l'évènement redouté. Ces séquences peuvent être réparties parmi les trois grandes familles présentées au IV.A, qui illustrent chacune les différents contributeurs à une situation accidentelle.

Dans un premier temps, la génération de séquences s'est faite sur un modèle privé des évènements n'impactant que la production et n'ayant aucune conséquence sécuritaire, afin de

limiter l'explosion combinatoire. L'outil SimfiaNeo offre la possibilité d'inhiber une partie du code tout en conservant un modèle cohérent, afin de gérer les différentes configurations. Il est ainsi tout à fait possible d'utiliser le modèle complet pour rechercher des séquences menant à des incidents de production sans conséquences sécuritaires.

Ce calcul est globalement suffisant pour balayer les principales familles de scénarios accidentels car le modèle initial reste un modèle fonctionnel de haut niveau qui n'entre pas dans le détail des composants qui réalisent ces différentes fonctions. Néanmoins, une génération à l'ordre 3 serait insuffisante pour couvrir exhaustivement un modèle détaillé, menant à une troncature des résultats. Le niveau de granularité présenté par le modèle est toutefois pertinent pour identifier les séquences principales incluant l'évènement initiateur et les fonctions de sécurité mises en œuvre pour éviter d'atteindre la situation redoutée. Cette liste de séquences peut ainsi servir de point de départ à la construction des différents graphes, pour lesquels il reste à relier les composants réalisant les différentes fonctions de sécurité identifiées dans la séquence.

Comme décrit au IV.B, une séquence portant sur la possibilité d'obtenir un mélange explosif au sein d'un réservoir de stockage a été détaillée. Cette séquence a été initiée d'une part dans un graphe, complété par les composants réalisant les fonctions de sécurité mises en jeu et leurs liens respectifs. Le modèle MBSE-MBSA initial a été détaillé d'autre part pour intégrer ces composants, leurs défaillances et leur logique fonctionnelle, afin de pouvoir calculer en parallèle leurs facteurs d'importance. Il est ainsi possible de comparer les résultats des calculs de facteurs d'importance avec le calcul des indicateurs de centralité du graphe construit, dans le but d'identifier les composants mis en avant par les différentes méthodes.

Le calcul des facteurs d'importance a été réalisé en s'appuyant sur des données de fiabilité issues de la base OREDA [13] qui référence de nombreux équipements utilisés dans le domaine de l'*oil & gas*. Bien que ces équipements ne soient pas strictement identiques à ceux utilisés sur une installation de production d'hydrogène, nous nous intéressons avant tout à un ordre de grandeur réaliste entre les différents types de composants.

Ce calcul identifie parmi les éléments les plus critiques le compresseur permettant d'alimenter l'instrumentation en air comprimé pour la manœuvre des vannes. Il identifie également la vanne de mise à l'évent permettant le rejet du mélange hydrogène-oxygène lorsque l'analyseur détecte que la concentration d'oxygène dans le mélange s'approche trop de la limite inférieure d'explosivité (LIE). Le compresseur est notamment identifié comme l'un des principaux contributeurs aux pertes du fait de sa fiabilité inférieure aux autres équipements.

Le calcul des indicateurs de centralité met en évidence d'autres composants que sont l'automate de sécurité et les éléments de disjonction qui lui sont associés. Cet automate collecte les alarmes déclenchées par les boucles de sécurité surveillant les principaux paramètres de l'électrolyseur que sont la température, le débit d'eau entrant et sa conductivité. Un écart sur les valeurs de consigne est susceptible d'entrainer une déformation progressive de la membrane. L'automate apparait ainsi comme un élément central du réseau, à l'inverse du compresseur qui est situé sur une extrémité.

VI. Conclusion

Le projet P20-1 a permis l'exploration de nouvelles approches de modélisation des systèmes complexes, basées sur les jumeaux numériques, la théorie des réseaux ou les sciences du vivant. Un cas d'usage a été développé pour tester ces approches dans le cadre du *benchmark*. Ce cas d'usage repose sur une installation industrielle de production d'hydrogène. Cette installation est soumise à des enjeux de fiabilité et de disponibilité de production, de sûreté de fonctionnement au travers des dossiers réglementaires d'exploitation ainsi qu'au risque de cyber-malveillance.

Cette installation a été modélisée dans un premier temps par un modèle AltaRica. L'expressivité de ce langage formel est suffisante pour permettre la prise en compte de tous les enjeux et définir la sémantique utilisée au sein du modèle. Ce modèle sert ensuite de base commune pour explorer d'autres méthodes de traitement des systèmes complexes.

Les algorithmes d'exploitation du langage AltaRica permettent toutefois d'effectuer une première analyse et de calculer des indicateurs similaires à ceux obtenus lors des études classiques. Ainsi, la disponibilité de production est évaluée par un calcul stochastique tandis que la génération de séquences permet d'obtenir l'ensemble des séquences d'accident menant à un évènement redouté et de les quantifier. Ces séquences tirent profit du caractère unificateur du modèle, permettant de faire apparaitre dans les mêmes séquences, des défaillances d'équipements, des défauts de contrôle, des modifications de l'environnement ou des attaques cyber.

Le modèle AltaRica est construit sur un formalisme évènementiel discret qui permet de définir une sémantique capturant les enjeux de chaque discipline. Cependant, ce formalisme n'est pas assez représentatif pour restituer la finesse des phénomènes physiques observés. D'autre part, les indicateurs calculés par le modèle se différencient peu des indicateurs habituels. Les algorithmes d'exploitation utilisés présentent également le défaut d'avoir une complexité exponentielle, ce qui oblige régulièrement à tronquer les résultats pour trouver un bon compromis entre la représentativité et le temps de calcul. Il est donc nécessaire d'affiner ce modèle en essayant d'autres méthodes de traitement.

Une première option consiste à réaliser une analyse topologique du réseau obtenu en transformant les séquences accidentelles. Chaque séquence est transformée en un réseau partant de l'évènement initiateur jusqu'à l'évènement redouté et explorant les alternatives à chaque transition. Les indicateurs de centralité sont calculés sur ce réseau afin de déterminer de nouveaux facteurs d'importance permettant d'identifier des composants critiques, y compris ceux apparaissant dans des séquences rares potentiellement oubliées par les approches classiques ou les modèles simplificateurs. Cette approche sera cependant plus intéressante à évaluer sur un système de taille plus importante comportant plus d'interactions que le cas d'usage retenu.

Une seconde option consiste à s'inspirer des modèles construits pour les sciences du vivant. Les systèmes vivants interagissent à plusieurs niveaux et réagissent à leur environnement immédiat sans supervision centralisée. Il s'applique d'autres paradigmes de modélisation qui peuvent permettre d'identifier des propriétés différentes de celles qui sont habituellement évaluées sur les systèmes industriels. Ce champ nécessite néanmoins plus d'investigations pour réussir

à construire et exploiter les modèles sur un système industriel. Les outils disponibles sont pour l'instant difficiles à transposer hors de leur domaine de définition et ne couvrent qu'une petite partie de la complexité d'un système vivant. L'exploration des techniques de modélisation du vivant s'est faite en parallèle de la réalisation du cas d'usage, néanmoins sans application directe à celui-ci.

Ce travail a consisté en une première phase d'état de l'art permettant de recenser de nouvelles méthodes de traitement des systèmes complexes apparues depuis le projet P11-4 en 2012. Dans un second temps, certaines de ces méthodes ont été testées au travers d'un cas d'usage dont les principaux résultats sont détaillés dans cet article. En parallèle, ce travail d'exploration et de questionnement a permis d'ouvrir des perspectives intéressantes à approfondir dans des travaux futurs pour fertiliser le croisement des disciplines. Parmi ces perspectives figure la possibilité d'utiliser des réseaux multiplexes afin de pouvoir modéliser des phénomènes qui se superposent ou plusieurs scénarios issus du même initiateur, afin de pouvoir notamment superposer les perspectives de plusieurs disciplines au sein d'un même réseau.

REMERCIEMENTS

Les auteurs remercient tout particulièrement Daniel Krob, président du CESAMES et Véronique Thomas-Vaslin, immunologiste et chercheur au CNRS pour leur contribution à la réflexion autour des méthodes d'analyse des systèmes complexes. Les auteurs remercient également les souscripteurs pour le support technique apporté durant le projet, ainsi que l'IMdR pour l'organisation du projet P20-1.

REFERENCES

[1] IMdR Projet P11-4 « Etat de l'art des méthodes et outils innovants pour la modélisation des systèmes complexes », 2012

[2] Présentations faites à la manifestation IMdR : « Approches innovantes pour la maîtrise des systèmes complexes » Octobre 2009 et à la plateforme d'échange IMdR –2010

[3] Le Moigne J.L. La modélisation des systèmes complexes. Dunod editions, 1990

[4] Thèse de Mouna Rifi, «Modélisation et Analyse des Réseaux Complexes : application à la sûreté nucléaire», soutenue en 2019, (LIPN, Université Paris 13)

[5] C. Duval, G. Fallet-Fidry, B. Iung, P. Weber and E. Levrat, "A Bayesian network-based integrated risk analysis approach for industrial systems: application to heat sink system and prospects development", Journal of Risk and Reliability, 2012

[6] https://dcbrain.com/, Optimisation de réseaux Utilities & Supply Chain grâce à l'IA, 2021

[7] H. Zwirn. Les systèmes complexes – Mathématiques et biologie, Odile Jacob Sciences, 2006.

[8] Thomas-Vaslin, V. (2016). Contribution à une évaluation globale des systèmes complexes et des perturbations: l'exemple du système immunitaire (Doctoral dissertation, Ministère de l'Ecologie, du Développement durable et de l'Energie depuis 2010).

[9] Xavier de Bossoreille, Mathilde Machin, Laurent Sagaspe. Un nouvel outil de safety pour maîtriser la complexité des systèmes. Congrès Lambda Mu 21, " Maîtrise des risques et transformation numérique : opportunités et menaces ", Oct 2018, Reims, France.

[10] André Arnold, Alain Griffault, Gérald Point, and Antoine Rauzy, The AltaRica Formalism for Describing Concurrent Systems, In Fundamenta Informaticae. IOS Press. Vol. 34, pp 109–124, 2000.

[11] Frédérique Bicking, Christophe Simon, Walid Mechri. Mesures d'importance par réseaux bayésiens. 10ème Congrès International Pluridisciplinaire Qualité et Sûreté de Fonctionnement, Qualita'2013, Mar 2013, Compiègne, France.

[12] Nacim Fateh Chikhi. Calcul de centralité et identification de structures de communautés dans les graphes de documents. Interface homme-machine [cs.HC]. Université Paul Sabatier - Toulouse III, 2010.

[13] OREDA Handbook 6th Edition, Volume I Topside equipment, 2015