université PARIS-SACLAY

NNT : 2019SACLN049

Thèse de doctorat

Apport des paradigmes des Systèmes à Evènements Discrets pour la réduction du flux d'alarmes industrielles

Thèse de doctorat de l'Université Paris-Saclay

préparée à l'Ecole normale supérieure Paris-Saclay

École doctorale n°580 Sciences et Technologies de l'Information et de la Communication (STIC) spécialité Automatique.

Thèse présentée et soutenue à Cachan, le 28 novembre 2019, par

Yannick Laumonier

Composition du Jury :

Bruno Monsuez	
Professeur des universités, ENSTA ParisTech	Président
Jean-François Pétin	
Professeur des universités, Université de Lorraine	Rapporteur
Audine Subias	
Maître de conférences HdR, INSA Toulouse	Rapporteur
Alexandre Philippot	
Maître de conférences HdR, Université de Reims Champagne-Ardenne	Examinateur
Jean-Marc Faure	
Professeur des universités, Ecole normale supérieure Paris-Saclay	Co-Directeur de thèse
Jean-Jacques Lesage	
Professeur des universités, Ecole normale supérieure Paris-Saclay	Co-Directeur de thèse
Hervé Sabot	
Engineering director, General Electric Digital Europe	Invité

Remerciements

Tout d'abord, merci à Jean-Jacques Lesage et Jean-Marc Faure de m'avoir pris sous leurs ailes. Nos discussions et vos remarques m'ont permis de m'améliorer tout au long de cette thèse. Merci Jean-Jacques de m'avoir fait découvrir les Systèmes à Evènements Discrets durant mon cursus à l'ENS et de m'avoir guidé jusqu'au doctorat. Merci Jean-Marc d'avoir proposé ce sujet de thèse sur une problématique qui continue de m'intéresser après ces trois ans. Merci également de m'avoir permis de donner des cours à Supméca, cela m'a fait découvrir une partie du métier d'enseignant.

Je tiens également à remercier Hervé Sabot qui m'a offert la possibilité de réaliser ma thèse en immersion chez General Electric. Sans cela, il n'aurait pas été possible d'adapter mes travaux aux besoins réels de l'industrie. Merci Hervé d'avoir toujours été présent lorsque des difficultés venaient ralentir la progression de la thèse, et de m'avoir permis d'adapter mon emploi du temps pour travailler aussi bien sur les besoins universitaires qu'industriels. Merci également à tous mes collègues de General Electric (actuels et anciens), tout particulièrement Suzy et Stéphane, pour la bonne ambiance et l'intérêt que vous avez porté à mes travaux.

Merci à tous les doctorants (et jeunes docteurs) du LURPA pour ces trois ans. Le CIVIL (et ses soirées) et les différentes traditions du laboratoire sont des piliers de la bonne ambiance et de l'émulation positive qui règnent au laboratoire. Merci Kévin d'avoir su partager un bureau et un directeur de thèse, sans tes conseils et ton attitude positive je n'aurais sans doute pas aussi bien vécu la thèse. Merci Yann, tu m'as supporté dans le bureau un an de plus que Kévin et tu m'as montré les subtilités de la rédaction d'un manuscrit de thèse. Merci Florian pour m'avoir permis de relâcher la pression, ne change surtout pas. Merci Clément, le fait que nous ayons commencé la thèse en même temps et dans la même équipe m'a permis de relativiser lorsque j'avais l'impression que je n'avançais plus. De plus, tu as su nous faire décompresser en nous communiquant ta curiosité et en nous faisant participer à certains de tes projets.

Merci également à tous mes amis qui m'ont soutenu et encouragé durant ma thèse.

Merci à ma famille d'avoir toujours cru en moi et de m'avoir soutenu jusque-là. D'avoir été présente, physiquement et/ou mentalement à ma soutenance. Contrairement à ce que vous dîtes, je n'en saurais pas là sans vous.

Table des matières

Introduction générale 11

1) Motivations 11

2) Contributions 11

3) Organisation du manuscrit 12

Chapitre 1 : Gestion des alarmes industrielles 13

1) Les systèmes de gestion d'alarmes 14

a. Définition d'une alarme industrielle 14

b. Architecture d'un système moderne de contrôle et de supervision 16

c. Le cycle de vie d'un système de gestion d'alarmes 19

2) La complexification de la gestion d'alarmes 21

a. Surabondance d'alarmes 21

b. Criticité des systèmes d'alarmes : 24

3) Les propositions industrielles pour faciliter la gestion des alarmes 25

a. Améliorer l'interface Homme/machine 26

b. Analyser des historiques d'alarmes 28

c. Problématique industrielle de la thèse 30

4) Conclusion 32

Chapitre 2 : Etude bibliographique et positionnement des travaux 33

1) Etat de l'Art 34

a. Etudes n'utilisant pas d'historiques d'alarmes 35

b. Etudes basées sur l'analyse d'historique d'alarmes 37

2) Approche proposée 43

a. Méthode générale 43

b. Formulation grâce aux paradigmes des Systèmes à Evènements Discrets 44

Chapitre 3 : Identification d'adjacences fréquentes entre alarmes 51

1) Réduction de la séquence par analyse de motifs fréquents 52

a. Support d'un motif et motif fréquent 52

b. Obtention des séquences équivalentes 53
2) Algorithme d'identification des motifs fréquents 54
a. Choix de l'algorithme 54
b. Description de *AprioriAll* 55
c. Algorithme développé 59
3) Présentation de l'exemple 63
4) Application de la méthode sur l'exemple 64
a. Obtention des motifs fréquents 64
b. La détermination des séquences équivalentes 65
c. Analyse experte 66
5) Conclusions 67
Chapitre 4 : Identification de précédences systématiques entre alarmes 69
1) Réduction de la séquence par analyse de précédences systématiques 70
a. Analyse sur des séquences projetés 70
b. Obtenir la séquence équivalente 70
2) Obtenir les précédences systématiques 71
a. Relation de domination 71
b. Relation de mutuelle dépendance 72
3) Représenter les précédences systématiques 74
a. Traduction d'une domination 75
b. Traduction d'une mutuelle dépendance 76
c. Assemblage des fragments 76
d. Réduction des places implicites 78
4) Application au cas de la chaudière 79
a. Le réseau de Petri traduisant les précédences systématiques entre alarmes 79
b. Les séquences équivalentes à chaque fragment 80
c. L'analyse experte 82
5) Conclusions 84
Chapitre 5 : Traitement d'une étude de cas industrielle et comparaison des méthodes 87

1) La gestion industrielle des historiques d'alarmes 88
a. Le stockage et l'utilisation des données 88
b. Les données utilisées pour l'étude comparative 89
2) Application de la méthode de recherche d'adjacences fréquentes 90
a. Application sur un jeu de données industriel 90
b. Implantation dans la plateforme Predix 93
3) Application de la méthode de recherche de précédences systématiques 97
a. Application sur un jeu de données industriel 97
b. Implantation dans la plateforme Predix 101
4) Comparaison des deux méthodes 101
a. Comparaison numérique des résultats 101
b. Comparaison qualitative et experte des résultats 103
c. Résultats de la comparaison 106
5) Conclusions et discussion 107
Conclusions générales et perspectives 109
1) Résumé des travaux 109
2) Conclusions 110
3) Perspectives 110
a. L'utilisation d'informations contenues dans les historiques d'alarmes 110
b. L'utilisation de connaissances sur l'évolution du processus 111
Références bibliographiques 113

Table des Figures

Figure 1 : Exemple de définition d'une alarme à partir d'une mesure capteur (Wuhib et al. 2005) 15
Figure 2 : Architecture type d'un système moderne de contrôle et de supervision (source GE) 18
Figure 3 : Cycle de vie d'un système de gestion d'alarmes (International Society of Automation 2009) 19
Figure 4 : Évolution du nombre d'alarmes configurées en pétrochimie (source Petrovue) 23
Figure 5 : Photographie de l'incident Texaco 25
Figure 6 : Evolution de l'interface utilisateur associée à une turbine (source GE) 27
Figure 7 : Exemple d'extrait d'un historique d'alarme (source : GE) 30
Figure 8 : Graphe de densité de probabilités que les alarmes soient liées (issu de (Kinghorst et al. 2018)) 38
Figure 9 : Cartographie de la similarité entre alarmes (issu de (Kondaveeti et al. 2012)) 39
Figure 10 : Schéma détaillant l'approche proposée 43
Figure 11 : Exemple de configuration d'alarme 45
Figure 12 : Exemple de séquence d'alarmes 46
Figure 13 : Exemples de motifs 53
Figure 14 : Approches classiques pour fenêtrer une séquence (inspiré par (Krishnan, et al., 2014)) 59
Figure 15 : Exemples de fenêtrage avec une fenêtre de longueur 4 60
Figure 16 : Exemple simple d'application de l'algorithme AprioriAll 62
Figure 17 : Schéma représentant un système de production d'énergie (Mikkonen, et al., 2016) 63
Figure 18 : Exemple de domination dans une séquence d'alarmes 72
Figure 19 : Exemple de mutuelle dépendance dans une séquence d'alarmes 73
Figure 20 : Représentation d'une relation de domination 75
Figure 21 : Représentation d'une relation de mutuelle dépendance 76
Figure 22 : Assemblage des relations au niveau des transitions communes 77
Figure 23 : Représentation complète des relations entre alarmes 77
Figure 24 : Représentation complète des relations entre alarmes avec suppression des places implicites 79
Figure 25 : RdP traduisant les relations identifiées sur l'exemple de la chaudière 80
Figure 26 : Evolution du nombre d'occurrences d'alarmes durant le mois 89

Figure 27 : Evolution du nombre d'occurrences d'alarmes durant le mois avant et après la réduction des séquences fréquentes 92
Figure 28 : Rendu visuel de l'application 93
Figure 29 : Etapes d'initialisation de l'application 94
Figure 30 : Affichage de l'évolution temporelle du nombre d'occurrences d'alarmes 95
Figure 31 : Répartition des alarmes configurées en fonction de leur priorité 95
Figure 32 : Résultats de la recherche limitée de séquences fréquentes 96
Figure 33 : Exemples de fragments obtenus a) avec alarmes occurrant le plus souvent, b) avec alarmes occurrant le moins souvent 98
Figure 34 : Exemple de fragment composé de dominations et d'une mutuelle dépendance .. 98
Figure 35 : Evolution du nombre d'occurrences d'alarmes durant le mois avant et après la réduction permise par les fragments de RdP 100
Figure 36 : Mise en parallèle des résultats des méthodes par recherche a) d'adjacences fréquentes et b) de précédences systématiques 102
Figure 37 : Réseau de Petri identifié dans l'exemple de la chaudière 104
Figure 38 : Fragment avec informations utiles à l'expertise 107

Liste des tableaux

Tableau 1 : Étude multi-industrielle du nombre d'alarmes à gérer (EEMUA 2007) 24
Tableau 2 : Produits achetés par transaction et par client (inspiré de (Agrawal & Srikant 1995)) 56
Tableau 3: Séquences élémentaires (inspiré de (Agrawal & Srikant 1995)) 56
Tableau 4 : Suites d'achats (inspiré de (Agrawal & Srikant 1995)) 57
Tableau 5 : Séquences fréquentes de longueur 2 (inspiré de (Agrawal & Srikant 1995)) 57
Tableau 6 : Séquences fréquentes de longueur 3 (inspiré de (Agrawal & Srikant 1995)) 58
Tableau 7 : Séquences fréquentes de longueur 4 (inspiré de (Agrawal & Srikant 1995)) 58
Tableau 8 : Séquences fréquentes maximales (inspiré de (Agrawal & Srikant 1995)) 58
Tableau 9 : Motifs fréquents identifiés par AprioriAll 65
Tableau 10 : Séquences fréquentes identifiées 91

Introduction générale

1) Motivations

Dans le contexte actuel de mondialisation, les entreprises doivent redoubler d'efforts pour rester rentables. L'un des enjeux est alors d'améliorer la disponibilité de leurs processus industriels en évitant les dysfonctionnements, pannes et dérives. Les industriels cherchent donc à ajouter des alarmes pour augmenter le volume d'informations disponibles pour la conduite d'un processus et ainsi augmenter la sûreté de leurs systèmes.

Ces informations sont issues de capteurs installés sur les équipements industriels. Or, depuis 20 ans, les possibilités offertes par de nouvelles technologies de capteurs et, de manière concomitante la baisse des prix des capteurs, ont provoqué une hausse du nombre de capteurs utilisés. Grâce à tous les signaux générés par ces capteurs et à l'amélioration des logiciels de configuration d'alarmes, il est devenu très simple de créer des alarmes qui se déclenchent à chaque fois que le système sort de son domaine de fonctionnement désiré.

Cette facilité fait qu'aujourd'hui les systèmes de contrôle industriels souffrent d'un trop grand nombre d'alarmes qui gênent le travail des opérateurs en les inondant d'informations qu'ils n'ont pas le temps de traiter. Ce phénomène provoque de nombreux problèmes pouvant aller d'une simple baisse de disponibilité sur le système à sa destruction.

Cependant, les alarmes étant un moyen pour les opérateurs de surveiller et de contrôler le bon fonctionnement d'un système industriel, les supprimer indistinctement provoquerait une perte de sûreté. Seule une connaissance experte du système et de la configuration de ses alarmes permet d'identifier celles qui transportent des informations redondantes et dont la suppression ne provoque pas une perte de sûreté pour ce système.

Nos travaux visent à réduire le flux d'alarmes industrielles en associant une analyse automatique d'historiques d'alarmes à une expertise industrielle. Cette réduction consiste en un filtrage des alarmes redondantes dont la suppression n'entraîne pas une baisse de sûreté du processus surveillé.

2) Contributions

Pour mettre en place ce filtrage d'alarmes, nous proposons une approche qui utilise la connaissance du flux d'alarmes reçues par les opérateurs contenue dans des bases de données historiques. Grâce à cette connaissance, nous modélisons le flux d'alarmes comme une

séquence d'évènements. En considérant que si des alarmes sont redondantes, elles n'apparaissent pas dans un ordre aléatoire, il existe donc des relations entre leurs occurrences que nos travaux permettent de trouver de manière automatique. Ces relations permettent d'identifier des alarmes qui peuvent potentiellement être filtrées. Par la suite, un expert analyse ces résultats et choisit qu'elles sont les alarmes qui sont filtrées sans provoquer de perte de sûreté pour le processus industriel.

Pour identifier ces relations, nous proposons des méthodes formalisées détectant deux types de relations : les adjacences fréquentes (c'est-à-dire des alarmes apparaissant fréquemment côte à côte dans la séquence) et les précédences systématiques (c'est-à-dire des alarmes qui apparaissent toujours avant une autre alarme). Ces méthodes sont ensuite testées sur des données industrielles pour pouvoir quantifier leur apport réel à la problématique de réduction du flux d'alarmes.

3) Organisation du manuscrit

Le manuscrit est organisé comme suit. La manière de gérer des alarmes, les problèmes liés à cette gestion ainsi que des propositions industrielles sur la manière d'y remédier sont présentés dans le Chapitre 1. Cette présentation permet d'extraire la problématique qui motive ce travail d'un point de vue industriel. Par la suite, dans le Chapitre 2, nous développons un état de l'art des différentes contributions académiques qui cherchent à résoudre ce problème pour pouvoir positionner notre approche parmi ces différents travaux.

Dans notre approche, nous utilisons deux différentes méthodes formalisées pour identifier des relations entre alarmes. La première méthode consistant à identifier des adjacences fréquentes est développée dans le Chapitre 3, tandis que la seconde qui cherche des précédences systématiques est expliquée dans le Chapitre 4.

Elles sont toutes les deux confrontées à un jeu de données industriel pour les comparer et vérifier qu'elles offrent une solution acceptable au problème de la réduction du flux d'alarmes. Cette confrontation est détaillée dans le Chapitre 5. Ce chapitre indique également la manière dont nos méthodes ont été implémentées sur une plateforme logicielle utilisée dans un contexte industriel.

Finalement, un résumé des travaux effectués, des conclusions et différentes perspectives sont donnés.

Chapitre 1 : Gestion des alarmes industrielles

Dans ce premier chapitre, après avoir défini ce qu'est une alarme, nous présenterons la manière dont sont gérées les alarmes dans un système de contrôle et de supervision. Pour cela, nous nous attarderons sur la manière dont les alarmes sont configurées et traitées dans un contexte industriel en décrivant le cycle de vie d'un système de gestion d'alarmes. Par la suite, nous préciserons en quoi cette gestion est complexe et comment la majorité des entreprises industrielles gère cette difficulté. Enfin, nous définirons les problèmes sous-jacents à la complexification croissante des systèmes d'alarmes, ce qui nous conduira à la définition de la problématique industrielle qui est à l'origine de cette thèse.

1) Les systèmes de gestion d'alarmes

a. Définition d'une alarme industrielle

L'*American National Standards Institute* (ANSI) et l'*International Society of Automation* (ISA) ont collaboré pour définir la notion d'alarme, ainsi que la méthode de gestion des alarmes jugée optimale, en créant la norme ANSI/ISA 18.2 (International Society of Automation 2009).

D'après cette norme, une alarme est définie comme étant « An audible and/or visible means of indicating to the operator an equipment malfunction, process deviation or abnormal condition requiring a timely response ». C'est-à-dire qu'une alarme est un signal audible et/ou visible indiquant à un opérateur une défaillance d'un équipement, une déviation du processus ou toute condition anormale qui nécessite une réponse rapide de la part de cet opérateur.

Comme, par définition, chaque alarme nécessite une réponse d'un opérateur, il est important que ce dernier ait le temps d'analyser chaque alarme, ainsi que ses causes et ses conséquences, pour comprendre et résoudre le problème ayant généré cette alarme. Il est donc primordial que les opérateurs ne soient pas débordés par le nombre d'alarmes pour qu'ils aient le temps d'accomplir leur mission correctement. Ainsi, si aucune action de l'opérateur n'est nécessaire, le signal ne doit pas être déclaré en tant qu'alarme, mais en tant que simple évènement informatif de l'état du système. Il est donc nécessaire de s'interroger dès la création du système d'alarme sur la nature des informations transmises à l'opérateur et notamment sur la nécessité de les qualifier en tant qu'alarme.

Une alarme peut être créée à partir de mesures réalisées sur le système grâce à des capteurs, ou à partir de toute autre information sur l'état du processus. Un exemple d'alarme définie à partir d'une mesure réalisée par un capteur est illustré sur la Figure 1 :

- La date d'apparition (de début) d'une alarme est définie comme la première date où une mesure a atteint une valeur supérieure à une valeur seuil T_g^+ (rising trigger threshold) depuis la date où la mesure a dépassé une valeur seuil T_g^- (falling trigger threshold) ($T_g^- < T_g^+$).
- La date de disparition (de fin) d'une alarme est définie comme la première date où une mesure a atteint une valeur inférieure à une valeur seuil T_g^- depuis la date où la mesure a dépassé une valeur seuil T_g^+ ($T_g^- < T_g^+$).

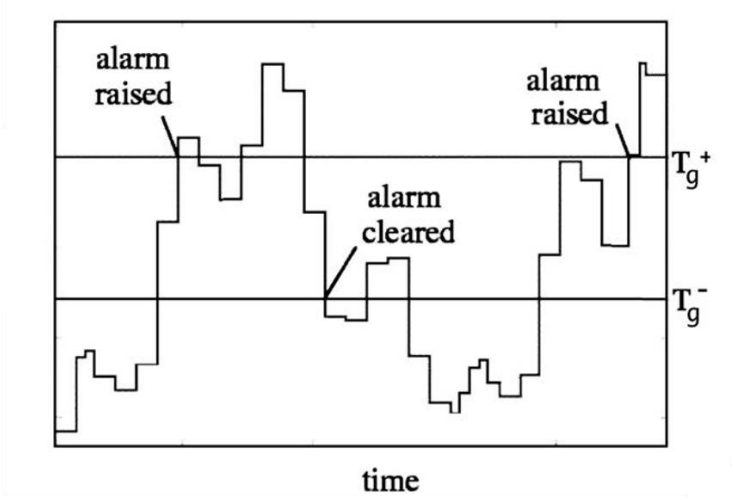


Figure 1 : Exemple de définition d'une alarme à partir d'une mesure capteur (Wuhib et al. 2005)

L'intérêt d'utiliser deux seuils T_g^+ et T_g^- à la place d'un seul seuil T_g^+ est que, si le signal oscille autour de T_g^+ à cause d'un bruit de mesure, plusieurs alarmes ne soient pas émises. L'utilisation de deux seuils correctement définis permet donc d'éviter les alarmes bruyantes (alarmes qui apparaissent plusieurs fois dans un court laps de temps).

Plusieurs alarmes peuvent être associées à une même mesure capteur en définissant des seuils de déclenchement différents. Par exemple, une mesure de pression qui dépasserait un seuil « haut » déclencherait une alarme « pression haute », mais déclencherait aussi une alarme « pression très haute» si elle dépasse un seuil « très haut ».

Une autre date qui est généralement utilisée lors de l'analyse du système d'alarme est la date d'acquittement d'une alarme. L'acquittement d'une alarme est une action de l'opérateur mettant fin au signal visuel ou sonore et indiquant que l'opérateur a bien reçu l'alarme. Après avoir acquitté l'alarme, il peut commencer à traiter l'alarme et résoudre le problème qui l'a déclenchée. Cependant, il est possible qu'il n'y ait pas d'action immédiate à effectuer en fonction de la nature du problème rencontré (Action de maintenance à étaler dans le temps, faible niveau d'urgence de l'alarme…). C'est le cas par exemple si une action corrective

effectuée pour une autre alarme fait également disparaître l'alarme considérée, ou si en évoluant naturellement le système retourne dans un état normal (cas d'une alarme transitoire). L'acquittement d'une alarme est donc bien différent de la disparition d'une alarme. Cependant, cet acquittement participe à la charge de l'opérateur et fait partie de la chaîne complexe d'actions à réaliser après ce type d'évènement.

Il y a donc 3 dates qu'il est intéressant d'observer : les dates d'apparition et de disparition (dates liées à des évènements générés par les processus) et la date d'acquittement (date générée par un évènement opérateur)

Ainsi, dans un exemple de scénario où une chaudière subit une surpression, l'alarme « Pression trop haute dans chaudière » apparaît. L'opérateur reçoit cette alarme et l'acquitte pour indiquer qu'il a bien vu l'alarme. Il cherche alors les raisons et les conséquences de cette alarme et décide des actions correctives à mettre en place. Dans ce scénario, l'opérateur décide par exemple d'ouvrir une vanne de surpression pour faire chuter la pression dans la chaudière. Une fois la pression redescendue à un niveau normal, l'alarme disparaît. Cet exemple, tout comme la Figure 1, illustre le cas d'une alarme basée sur une mesure capteur. Cependant, il existe d'autres types d'alarmes, tels qu'une alarme indiquant un dépassement de temps, ou encore une alarme associée à un dysfonctionnement du système d'acquisition.

Les alarmes, et en particulier les alarmes basées sur des mesures capteurs, utilisent des informations du système industriel pour permettre la prise de décisions correctives sur ce système. Les alarmes font donc partie intégrante du système de contrôle et de supervision, dont nous allons rappeler l'architecture type.

b. Architecture d'un système moderne de contrôle et de supervision

Un système moderne de contrôle et de supervision possède généralement une architecture similaire à celle de la pyramide de l'automatisation, mais avec quelques différences dues aux progrès des technologies numériques. De ce fait, comme l'indique la Figure 2 issue de documents General Electric, cette architecture repose toujours sur la notion de niveaux, chacun correspondant à un niveau d'automatisation du processus :

- Niveau 0 – Niveau processus : ce niveau correspond au processus industriel brut dépourvu d'automatisme;
- Niveau 1 – Niveau capteurs : ce niveau correspond aux capteurs et à tout ce qui se rapporte à la récupération d'informations sur l'état du processus ;

- Niveau 2 – Niveau de contrôle : ce niveau correspond aux automates programmables industriels (PLCs) et à tout ce qui permet d'agir sur le processus et de contrôler localement son évolution ;
- Niveau 3 – Niveau de supervision : ce niveau se compose de l'ensemble des équipements (matériels ou numériques) permettant de superviser et de piloter le comportement du processus. C'est également à ce niveau que les opérateurs en charge des alarmes se situent ;
- Niveau 4 – Niveau de planification : ce niveau correspond à la gestion des ressources et à la gestion temporelle des phases du processus telles que la programmation de phases de maintenance. Dans le cadre de l'industrie 4.0, ce niveau correspond également à la frontière entre le réseau de l'entreprise et le réseau Internet industriel. Il peut donc être séparé en deux sous-niveaux avec un niveau OT (operational technology), "interne" (le réseau de l'entreprise), et un niveau IT (information technology), "externe" (sur le réseau Internet).

La plateforme Predix sur laquelle nous implantons nos différents travaux est un environnement permettant de développer des applications industrielles et fait partie du niveau IT (nommé « Planning level (out) » sur la Figure 2).

La Figure 2 représente l'architecture type d'un système de contrôle et de supervision sur laquelle est représenté le chemin d'une mesure d'un capteur de température qui provoque une alarme traitée par un opérateur.

En plus de représenter les niveaux indiqués précédemment, la Figure 2 détaille les niveaux de supervision et de planification pour faire apparaître les différentes ressources de ces niveaux. Par la suite, nous retiendrons et détaillerons les ressources du niveau de supervision qui jouent un rôle prépondérant dans la gestion des alarmes industrielles.

Les informations concernant l'état du système remontent depuis le niveau des capteurs vers le niveau de supervision (en pointillés ronds roses sur la Figure 2). Ces informations sont gérées par un serveur SCADA (Supervisory Control And Data Acquisition) qui permet de traiter en temps réel un grand nombre de télémesures et de contrôler à distance des installations techniques. C'est ce serveur qui va décider si l'information correspond à un comportement anormal et si une alarme a besoin d'être générée et remontée à l'opérateur.

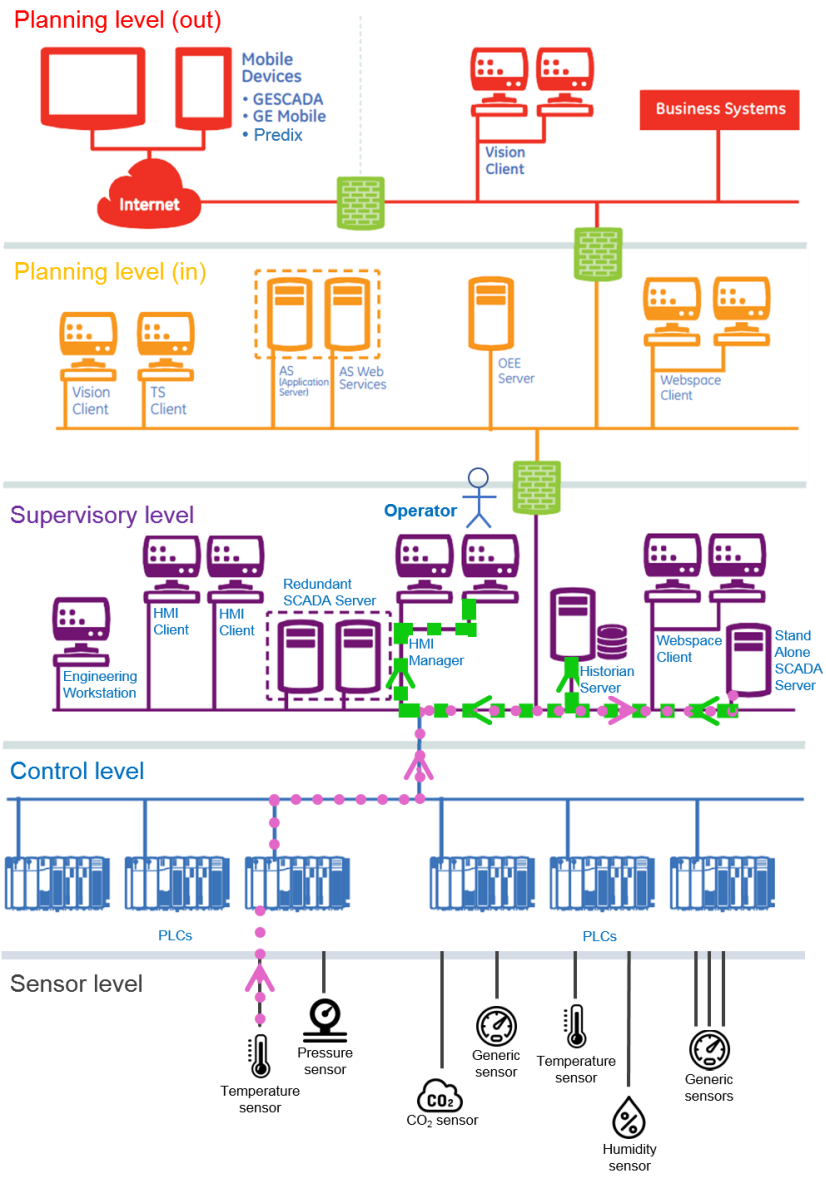


Figure 2 : Architecture type d'un système moderne de contrôle et de supervision (source GE)

Si l'information remontée correspond à une situation anormale nécessitant une action de la part d'un opérateur, une alarme est alors générée et transmise à un opérateur (trajet en pointillés carrés verts sur la Figure 2). Ce dernier est souvent référencé par son poste de contrôle qui sert d'interface Homme/machine. En plus d'être transmises à l'opérateur, l'alarme ainsi que les informations la concernant sont stockées dans une base de données génériquement appelée « Historian ». Cette base de données permet de conserver un historique de l'évolution du processus de manière à pouvoir analyser son comportement et l'améliorer.

Bien que le serveur SCADA analyse les informations venant des niveaux inférieurs automatiquement, c'est l'Homme qui doit traiter les alarmes générées par ce serveur. De ce

fait, pour faciliter le travail des opérateurs, un système de gestion d'alarmes a besoin de respecter des règles de bonnes pratiques. L'ensemble de ces règles se retrouve dans le cycle de vie des systèmes de gestion d'alarmes.

c. Le cycle de vie d'un système de gestion d'alarmes

Un bon système de gestion d'alarmes est indispensable pour assurer une bonne conduite du processus industriel. Cependant, celui-ci est configuré par des humains et il est difficile de s'assurer de son efficacité avant de le tester. De ce fait, un système de gestion d'alarmes doit faire l'objet d'une amélioration continue au fur et à mesure des itérations de son cycle de vie, tel que défini dans (International Society of Automation 2009) et présenté sur la Figure 3 :

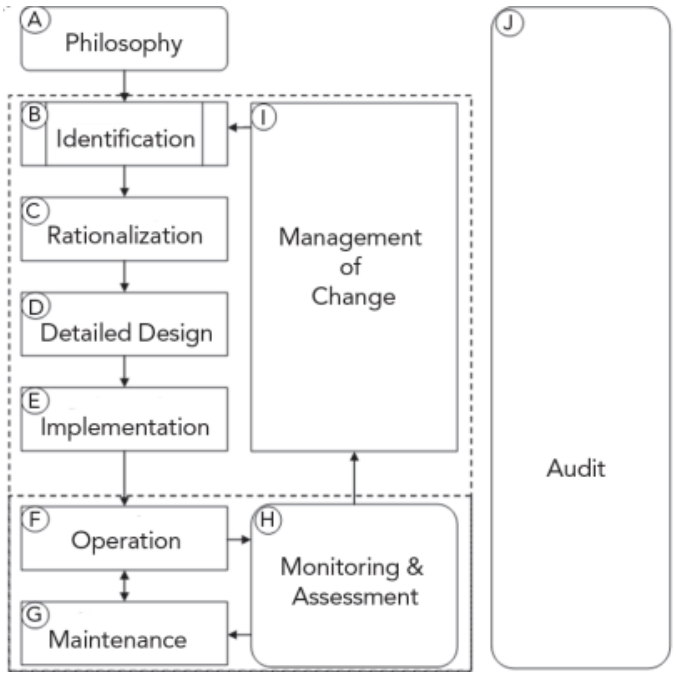


Figure 3 : Cycle de vie d'un système de gestion d'alarmes (International Society of Automation 2009)

Plusieurs personnes se partagent la responsabilité d'assurer le respect de ce cycle de vie. Ainsi, un responsable ayant une vision globale du système prendra en charge l'étape de définition de la philosophie. Un expert de ce même système aura pour tâche de réaliser les étapes d'identification, de rationalisation, de conception détaillée, d'implémentation et de gestion du changement. L'opérateur devra quant à lui gérer les phases de fonctionnement et de maintenance, mais il devra aussi surveiller et évaluer le système en proposant des modifications nécessaires à l'amélioration continue de ce système. Enfin, une entité externe indépendante aura pour mission de réaliser des audits pour vérifier l'intégrité du système de gestion d'alarmes et le respect du processus de gestion des modifications.

L'étape la plus importante qui va conditionner tout le fonctionnement du système de gestion d'alarmes est l'étape de définition de la philosophie du système (bloc A sur la Figure 3). La **philosophie** d'un système d'alarmes regroupe l'ensemble des objectifs de conception et d'exploitation du système avec les moyens à mettre en place pour atteindre ces objectifs. Une philosophie peut être commune à différents systèmes de gestion d'alarmes et peut se présenter sous la forme de différents principes tels que :

- « Les alarmes doivent être présentées à une vitesse telle qu'un opérateur puisse les gérer (en particulier quand le système industriel est perturbé ou dans des conditions anormales). »
- « Les alarmes doivent être faciles à comprendre. »
- « Les alarmes doivent être priorisées, le but étant d'aider l'opérateur à prendre des décisions quant à l'attention à porter à chaque alarme. »

La notion de priorisation des alarmes est très importante, et donner une signification sensée à chaque niveau d'alarmes est l'une des tâches essentielles pour définir la philosophie.

L'étape suivante consiste à déterminer quelles sont les alarmes nécessaires pour la bonne conduite du processus industriel. Cette étape s'appelle l'**identification** (bloc B sur la Figure 3), elle permet également de déterminer l'ordre dans lequel les alarmes doivent être traitées lors d'une nouvelle itération du cycle de vie (par exemple, celles qui apparaissent souvent).

Une fois les alarmes identifiées, il est nécessaire de s'assurer que ces alarmes respectent la philosophie définie auparavant. Cette démarche est la **rationalisation** (bloc C sur la Figure 3) du système de gestion d'alarmes.

Les définitions des alarmes étant justifiées et approuvées, il est possible de configurer les alarmes dans le système de gestion des alarmes. Cette étape, qui doit se faire dans le respect de la philosophie et de la phase de rationalisation, s'appelle la **conception détaillée** (bloc D sur la Figure 3).

Les alarmes désormais conçues sont mises en place de manière opérationnelle dans le système de gestion des alarmes et sont prêtes à être utilisées lors du fonctionnement du processus industriel. Cette étape d'**implémentation** (bloc E sur la Figure 3) est la dernière étape de création des alarmes.

Ces étapes sont du ressort d'un expert qui a pour fonction de configurer et d'implémenter les alarmes dans un serveur SCADA. Une fois les alarmes mises en place, celles-ci dépendent d'un opérateur qui doit les traiter au fur et à mesure qu'il les reçoit. Si le système de gestion

d'alarmes et l'équipement lié à l'alarme sont fonctionnels et que l'opérateur reçoit les alarmes, le système est alors dans sa phase **opérationnelle** (bloc F sur la Figure 3). Sinon, le système est dans une phase de **maintenance** (bloc G sur la Figure 3) qui doit avoir pour objectif de remédier aux différents problèmes pour que le système redevienne opérationnel.

En parallèle, que le système soit opérationnel ou en maintenance, toutes les données concernant le traitement des alarmes par les opérateurs sont récupérées et sont comparées aux objectifs de la philosophie par des experts. Cette phase de **surveillance et** d'**évaluation** (bloc H sur la Figure 3) permet aussi bien d'expliquer et de documenter des mises en maintenance imprévues du système, que d'identifier quels changements de configuration du système d'alarmes devraient être effectués pour se rapprocher d'un fonctionnement optimal au plus proche de la philosophie.

Ces changements sont nécessaires pour améliorer la gestion des alarmes. Cependant, ils ne doivent pas être réalisés n'importe comment. L'étape de **gestion des modifications** (bloc I sur la Figure 3) suit une démarche spécifique définie dans la philosophie du système de gestion d'alarmes. Celle-ci doit donc permettre de répondre à des interrogations que se poserait l'expert telles que : « Ai-je le droit de modifier cette alarme de cette manière ? » ou encore « Cette modification dépend-elle de mon autorité ou dois-je demander à quelqu'un d'autre ? ».

Il est clair que le système de gestion d'alarmes doit évoluer ou être modifié régulièrement. Cependant, l'étape de gestion des modifications peut nécessiter d'accéder à l'historique du système. C'est pour cette raison que les informations recueillies lors de la surveillance et l'évaluation du système doivent être stockées dans une base de données et être disponibles pour l'étude visant à identifier les modifications nécessaires.

Parmi ces différentes étapes, celles qui sont le plus chronophages pour l'expert lorsqu'il doit apporter des modifications à la configuration du système d'alarme sont les étapes de rationalisation et de gestion du changement, auxquelles nos travaux se proposent de contribuer.

2) La complexification de la gestion d'alarmes

a. Surabondance d'alarmes

Au cours des dernières années, de nouvelles technologies de capteurs apparaissent régulièrement sur le marché. Cette disponibilité de capteurs toujours plus performants à connexion sans-fil et à bas coût a incité les industriels à multiplier le nombre de capteurs sur

leurs équipements, augmentant ainsi le nombre de signaux remontant vers le niveau de supervision.

Ce nombre croissant de signaux disponibles, associé à la meilleure ergonomie des logiciels de CAO d'automatisation, permet aux experts de configurer très facilement un très grand nombre d'alarmes. Or, plus il y a d'alarmes configurées dans un système industriel, plus les perturbations sont susceptibles d'être détectées, même celles n'ayant aucun effet néfaste sur le système. De ce fait, augmenter le nombre d'alarmes configurées augmente invariablement le nombre d'alarmes que les opérateurs auront à traiter. De plus, un trop grand nombre d'alarmes configurées dans le système rend plus lentes les étapes de rationalisation et de gestion du changement réalisées par l'expert. Globalement, augmenter de manière inconsidérée le nombre d'alarmes d'un système industriel augmente souvent inutilement la charge de travail des opérateurs et des experts.

Cette modernisation des systèmes industriels et la facilité de créer des alarmes qui en découle, cause aussi l'effet pervers de diminuer le temps que passent les experts sur la configuration initiale d'une alarme. En passant moins de temps lors de la configuration initiale, les experts peuvent commettre des erreurs de conception qui éloignent le système d'alarme de sa philosophie et qui peuvent créer des dépendances et des redondances entre alarmes.

Les redondances les plus courantes peuvent par exemple avoir les causes suivantes :

- Des liens physiques entre équipements : deux alarmes peuvent être redondantes car les équipements qui leur sont associés sont liés physiquement par le processus (par exemple, une température trop haute d'une chaudière (alarme 1) qui provoque une pression trop élevée dans sa conduite de sortie (alarme 2)).
- Des liens dans la chaîne numérique : deux alarmes peuvent être redondantes car leur lien dans la chaîne numérique n'a pas été identifié lors de leur création. Cela peut être par exemple, une alarme associée à un seuil de dépassement sur la mesure absolue d'un capteur (valeur numérique variant dans la plage de mesure permise par le capteur) et une autre alarme associée à un seuil sur la valeur relative de cette même mesure (pourcentage de la valeur maximale permise par la plage de mesure du capteur).
- Une confusion cause/conséquence : deux alarmes peuvent être redondantes car elles possèdent une cause commune et cette cause n'a pas d'alarme associée (par exemple, une vanne qui ne s'ouvre pas (situation anormale qui n'a pas été associée

à une alarme par erreur de conception), qui est la cause d'alarmes indiquant un débit faible en début et en fin d'une canalisation).

- Un effet psychologique : la crainte de manquer l'observation d'une évolution anormale du processus peut provoquer la multiplication d'alarmes qui ne sont pas nécessaires.

L'augmentation du nombre d'alarmes configurées dans un système de gestion d'alarmes est illustrée pour le domaine pétrochimique par la Figure 4 :

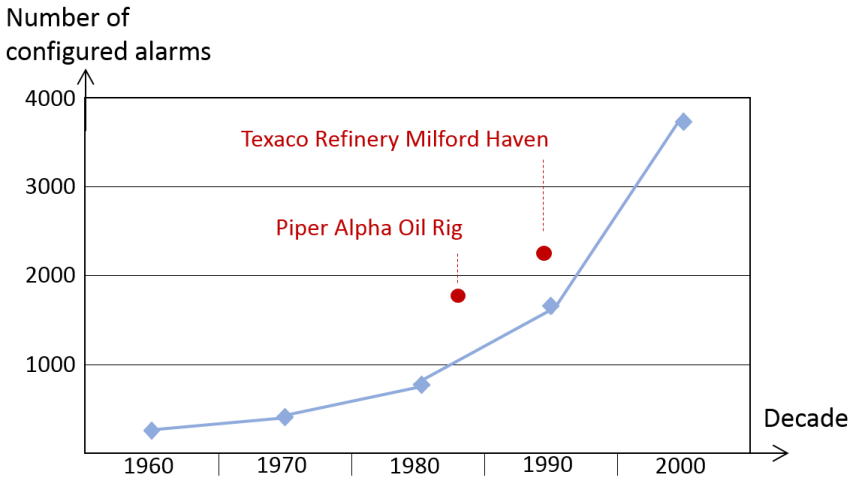


Figure 4 : Évolution du nombre d'alarmes configurées en pétrochimie (source Petrovue)

Sur cette figure, les points indiquent des catastrophes industrielles dont l'une des causes principales, d'après le rapport d'incident, était la mauvaise gestion des alarmes. L'incident catastrophique le plus représentatif et le plus célèbre quand on parle de mauvaise gestion d'alarmes est celui de la raffinerie de Texaco à Milford Haven. Cette catastrophe sera évoquée plus en détail par la suite.

Les points losanges indiquent pour chaque décennies la valeur moyenne du nombre d'alarmes configurées par les experts dans des processus industriels pétrochimiques. La tendance de la courbe associée à ces points montre bien l'effet de l'évolution des technologies de capteurs et des logiciels de CAO d'automatismes sur le nombre d'alarmes configurées. Toutes ces alarmes peuvent provoquer un grand volume d'informations à traiter et ainsi une surcharge de travail pour les opérateurs.

Les problèmes dus à cette surabondance d'alarmes se retrouvent dans tous les domaines industriels. Il a donc fallu créer des métriques permettant de quantifier la quantité d'alarmes

à gérer ainsi que la difficulté des opérateurs à traiter ces alarmes. Ainsi, les métriques retenues par l'*Engineering Equipment and Materials Users Association* (EEMUA 2007) sont : le nombre d'apparitions d'alarmes moyen par jour et le maximum d'apparition d'alarmes en 10 minutes. Les valeurs recommandées par l'EEMUA pour ces métriques dans le cas d'une bonne gestion des alarmes sont présentées dans le Tableau 1 et sont comparées aux valeurs communément observées dans différents domaines industriels :

Tableau 1 : Étude multi-industrielle du nombre d'alarmes à gérer (EEMUA 2007)

	Recommended by EEMUA	Observed in:		
		Oil-Gas	PetroChem	Power
Average alarms/day	144	1200	1500	2000
Peak alarms/10 min	10	220	180	350

Comme il est possible de le voir sur le Tableau 1, quel que soit le domaine industriel, les opérateurs ont beaucoup trop d'alarmes à gérer comparativement aux valeurs recommandées par l'EEMUA. Ces valeurs font office de référence comme objectif à atteindre pour que les opérateurs aient le temps d'analyser correctement, de comprendre et de corriger les problèmes associés à chacune des alarmes.

Globalement, ces mauvaises performances des systèmes de gestion d'alarmes sont d'autant plus problématiques que ces systèmes jouent un rôle critique dans la maîtrise du processus industriel considéré.

b. Criticité des systèmes d'alarmes :

Si l'une des métriques utilisées par l'EEMUA est le nombre maximum d'alarmes en 10 minutes, c'est qu'il existe un phénomène appelé **avalanche d'alarmes** qui est défini par ANSI/ISA 18.2 comme le fait d'avoir plus de 10 alarmes à traiter en 10 minutes. L'avalanche d'alarmes cesse quand l'opérateur doit de nouveau traiter moins de 5 alarmes en 10 minutes.

Une avalanche d'alarmes a des conséquences pouvant aller d'une légère diminution des performances à une catastrophe dramatique. Ainsi, dans le meilleur des cas, une avalanche d'alarmes va seulement provoquer une baisse de performance du système en diminuant sa disponibilité (à cause de machines rendues non opérationnelles) ou en augmentant les coûts de production.

Cependant, dans le pire des cas, une avalanche d'alarmes peut mener à la destruction du système et à de possibles pertes humaines. L'exemple le plus célèbre est celui de l'explosion de la raffinerie Texaco à Milford Haven, en 1994, montré sur la Figure 5.



Figure 5 : Photographie de l'incident Texaco

D'après le rapport d'incident (Health and Safety Executive 1997), la foudre en tombant a perturbé le fonctionnement de l'usine. Le rapport insiste cependant sur le fait que la foudre n'est pas la cause de l'explosion qui a eu lieu 5 heures plus tard. L'explosion n'est due qu'au fait que les opérateurs n'ont pas réussi à contrôler les effets de cette perturbation. Ainsi, le rapport met en cause la quantité d'alarmes que les opérateurs ont eu à gérer jusqu'à ce qu'ils soient totalement débordés. En effet, les opérateurs recevaient une alarme toutes les 2 ou 3 secondes durant l'incident et ont dû subir 275 alarmes durant les 10,7 minutes précédant l'explosion. Ces valeurs observées durant l'incident de Texaco sont aujourd'hui communes lors d'une avalanche d'alarmes dans le domaine de la production d'électricité (hors nucléaire).

Même si, grâce à des protections se déclenchant de manière réflexe, la destruction du système industriel due à une avalanche d'alarmes est devenue peu probable, il n'est économiquement pas possible de tolérer une mise à l'arrêt total trop fréquente d'une usine. C'est pour cela qu'il est important d'avoir un système d'alarmes bien configuré et qui permette aux opérateurs de se concentrer sur les alarmes les plus importantes. C'est pourquoi, les industriels ont développé et continuent de développer des solutions pour faciliter la gestion des alarmes.

3) Les propositions industrielles pour faciliter la gestion des alarmes

Le cycle de vie défini par ANSI/ISA 18.2 a été développé pour assurer une bonne gestion des alarmes (aussi bien lors de la configuration que lors de l'exploitation). En particulier, l'étape de rationalisation qui consiste à analyser les alarmes pour vérifier qu'elles respectent

la philosophie du système d'alarmes nécessite d'avoir une vision à la fois globale et locale de chaque alarme dans le cadre du processus industriel. Des solutions ont donc été mises en place pour obtenir cette vision. Ces dernières sont basées sur l'amélioration de l'interface Homme/machine et sur l'analyse des historiques d'alarmes.

a. Améliorer l'interface Homme/machine

Pour faciliter la gestion des alarmes, la solution choisie par la majorité des entreprises a été d'améliorer l'interface entre l'Homme (opérateur ou expert) et la machine. En plus d'aider la phase de rationalisation en visualisant l'intégration de l'alarme dans son environnement, cette approche présente également l'avantage d'aider directement l'opérateur dans sa prise de décision et d'améliorer sa vitesse de réaction. Ainsi, l'opérateur aura une interface intégrant la vision du processus en plus de l'interface dans laquelle les alarmes apparaissent et pourra rapidement visualiser le comportement du système pour chaque nouvelle alarme qu'il reçoit.

De manière générale, les informations utilisées par l'expert durant la phase de rationalisation et par l'opérateur durant la phase opérationnelle sont représentées sur une visualisation du système. Ces informations sont toutes celles qui peuvent présenter un intérêt pour comprendre la signification des alarmes ainsi que pour identifier leurs causes et leurs conséquences :

- L'équipement lié à cette alarme : le type d'équipement, la nature du capteur, la localisation de l'équipement dans le processus industriel ;
- La définition de l'alarme : une description de l'alarme, sa condition de déclenchement, la valeur du seuil de l'alarme (si alarme associée à une mesure capteur), le niveau de criticité de l'alarme ;
- L'environnement de l'alarme : les liens numériques (qui sont dues à la configuration logicielle des alarmes) entre l'alarme et toutes les autres variables ou alarmes s'ils existent, les liens physiques entre l'équipement associé à l'alarme et le reste du process ;
- Toutes autres informations jugées utiles et demandées par la philosophie du système d'alarmes.

Ainsi, (Koene & Vedam 2000) présente une solution d'interface développée par Honeywell. Il est cependant expliqué que cette solution n'est pas suffisante et qu'elle a besoin d'être améliorée pour éviter que les systèmes industriels continuent à être limités par le problème de la surabondance d'alarmes.

Bien qu'il soit admis que cette solution n'est pas suffisante, elle permet de grandement augmenter l'efficacité de l'opérateur à traiter les alarmes. C'est pourquoi, comme le montre la Figure 6, les interfaces Homme/machine continuent d'être améliorées au fil des années.

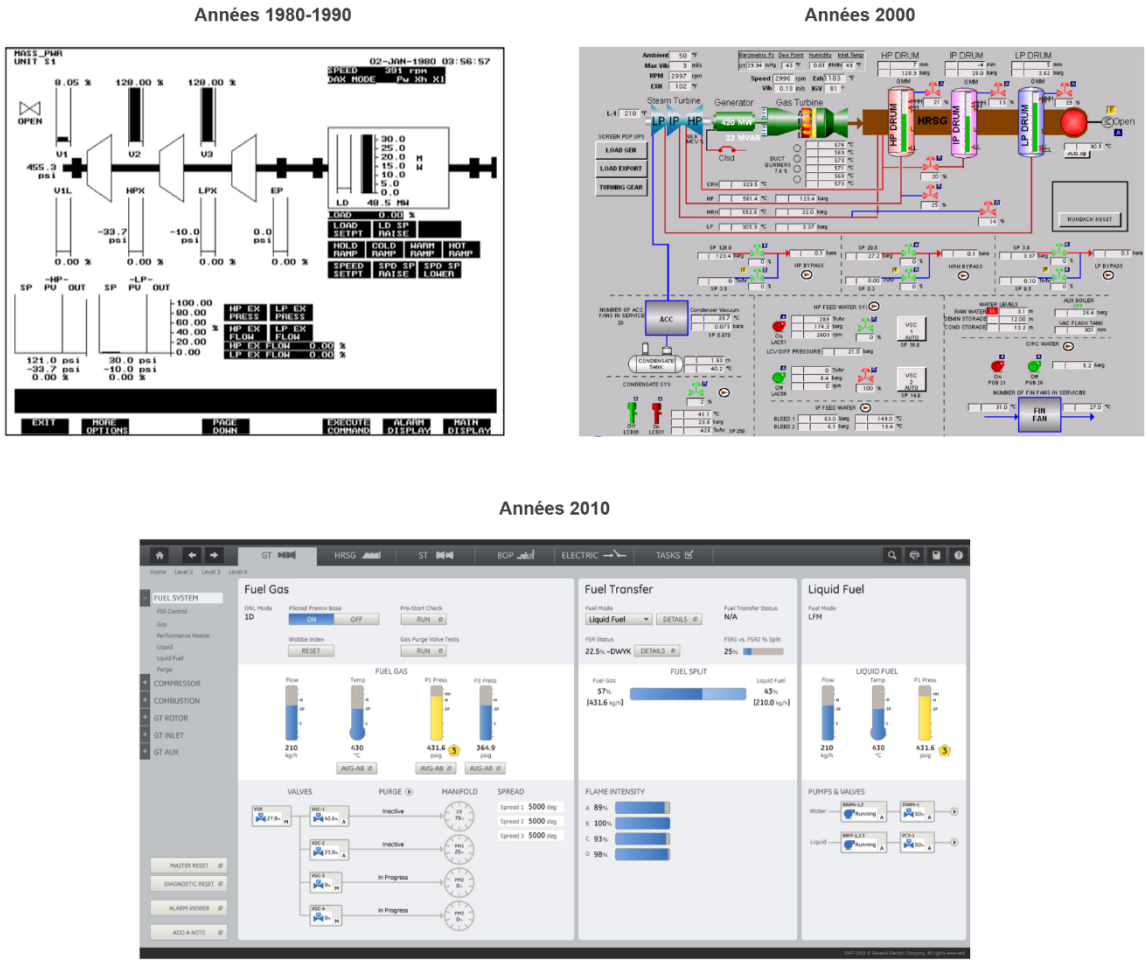


Figure 6 : Evolution de l'interface utilisateur associée à une turbine (source GE)

Sur la Figure 6, l'évolution d'une interface Homme/machine pour la surveillance d'une turbine General Electric est représentée. Il est possible de voir que sur chacune des interfaces, une représentation plus ou moins symbolique de la turbine est au cœur de l'interface (représentation complète pour les deux premières interfaces et fragmentée pour la plus récente). Sont également affichées des valeurs mesurées par des capteurs et des valeurs de variables numériques.

L'interface des années 1980-1990 est limitée par la technologie de son époque et affiche en noir et blanc une quantité limitée d'informations qui rend l'analyse des alarmes longue et difficile.

L'interface des années 2000 est à l'opposé de la précédente. Elle utilise une grande palette de couleurs pour catégoriser les informations et affiche beaucoup d'informations à l'écran. Ceci a pour effet d'augmenter le nombre d'informations que l'utilisateur doit surveiller et prendre en compte, ce qui est défavorable à sa vitesse de réaction.

L'interface des années 2010 tire parti des erreurs de celle des années 2000 et choisit de limiter le nombre d'informations présentées à l'utilisateur, tout en réduisant sa palette de couleurs pour augmenter le sens donné à chaque couleur. La turbine n'est plus affichée en entier, mais elle est séparée en sous-parties pour mieux cibler le problème à la source d'une alarme. Malgré cette séparation, l'utilisateur peut toujours visualiser les liens entre les parties en cliquant sur les variables ou les alarmes pour ouvrir un nouvel onglet contenant les informations complémentaires pertinentes à l'analyse. Cette interface épurée et hiérarchisée permet donc une meilleure détection de l'état du système et une analyse des alarmes plus rapide.

Malgré les progrès effectués sur l'interface Homme/machine (IHM) pour diminuer le temps de réponse d'un opérateur, le problème de la gestion d'un grand volume d'alarme persiste. En effet, l'analyse ne se faisant que sur une alarme à la fois, les opérateurs devront quand même éventuellement subir les effets d'une avalanche d'alarmes.

Les experts devant eux aussi analyser les alarmes une à une lors de la phase de rationalisation, cette rationalisation de l'IHM n'est pas suffisante pour les experts quand il y a un trop grand nombre d'alarmes configurées dans le système. Il faut donc apporter d'autres techniques qui permettraient d'identifier les alarmes ou groupements d'alarmes qui ont besoin d'une analyse approfondie de l'expert lors de la phase de rationalisation. Cette analyse plus poussée n'a pas forcément besoin de prendre en compte les alarmes en temps réel. Elle peut être effectuée durant la phase de gestion du changement puisque l'expert à l'avantage de pouvoir travailler en hors-ligne, en marge du processus industriel.

b. Analyser des historiques d'alarmes

Les analyses complémentaires à l'utilisation de l'interface Homme/machine pouvant être réalisées durant la phase de gestion du changement, les données à disposition pour l'analyse ne sont pas restreintes au flux d'alarmes en temps réel. Les données utilisées peuvent donc être celles recueillies durant la phase de surveillance et d'évaluation du cycle de vie. Elles contiennent les informations sur l'évolution passée des alarmes et forment un historique d'alarmes. Ces derniers représentent souvent plusieurs mois ou plusieurs années de fonctionnement du système étudié.

L'utilisation d'historiques d'alarmes pour effectuer des analyses utiles au travail de rationalisation d'un expert est une approche détaillée dans (Cabadi & Sabot 2016). Ce brevet défend le fait d'appliquer des algorithmes sur les historiques d'alarmes pour obtenir des informations complexes utiles pour l'expert.

Le besoin de conserver une trace de ce qu'il se passe lors de l'évolution d'un processus industriel est commun à différentes industries. C'est pour cela qu'une solution générique sous la forme d'un serveur appelé « historian » a été mise en place et est utilisée par la grande majorité des entreprises industrielles.

Comme évoqué précédemment, un serveur historian est à l'origine une base de données temporelles stockant les données relatives au contrôle et également à la supervision. Ces serveurs peuvent supporter des applications utilisant des algorithmes pour améliorer le contrôle qualité, le suivi des performances ou encore la supervision.

Un logiciel historian va donc stocker dans un serveur toutes les données qui pourraient être utiles à une future analyse. Dans le cadre d'une étude ayant pour but de rationaliser des alarmes, les informations stockées sont toutes celles qui ont un lien avec les alarmes et leur évolution. Pour chaque alarme, nous retrouvons les informations suivantes :

- L'identifiant de l'alarme ;
- La date d'apparition de l'alarme ;
- La date d'acquittement de l'alarme ;
- La date de disparition de l'alarme ;
- Le niveau de criticité de l'alarme ;
- Une description succincte de l'alarme ;
- L'équipement associé à l'alarme ;
- L'opérateur qui a géré l'alarme ;
- …

Pour illustrer ces propos, la Figure 7 présente un exemple d'historique d'alarmes que l'on peut trouver dans l'industrie du traitement de minerai. Dans cet extrait, « G » (pour « generated ») signifie nouvelle alarme générée et « R » (pour « reset ») signifie que l'alarme disparaît. De plus, « LOW », « MED » et « HIGH » sont les niveaux de criticité des alarmes.

timestamp	alarm_id	alarm_class	log_action	alarm_message
2014-02-08 14:13:12.677	alarm12	LOW	R	FLTR PRESS DISCH HOPPER EAST: Level High Alarm
2014-02-08 14:13:12.733	alarm6	LOW	R	FILTRATE DISCHARGE FLOW DENSITY: Level Open Wire Alarm
2014-02-08 14:13:12.777	alarm23	MED	R	FILTRATE DISCHARGE FLOW DENSITY: Level Low Alarm
2014-02-08 14:15:48.653	alarm17	LOW	G	EXCESSIVE CYCLE TIME: ALARM Alarm
2014-02-08 14:17:53.147	alarm8	HIGH	G	SD1 - FEED SOLIDS LOW WARNING: Density Low Alarm
2014-02-08 14:19:29.233	alarm8	HIGH	R	SD1 - FEED SOLIDS LOW WARNING: Density Low Alarm
2014-02-08 15:10:47.521	alarm4	LOW	G	PAL B - SAFETY PHOTO-CELL/SAFETY DOOR: Photo-cell/door Al
2014-02-08 15:10:47.623	alarm20	LOW	G	PAL B - EMERGENCY STOP: E-stop Alarm
2014-02-08 15:14:35.673	alarm13	LOW	G	PAL C - EMERGENCY STOP: E-stop Alarm

Figure 7 : Exemple d'extrait d'un historique d'alarme (source : GE)

Dans l'exemple présenté sur la Figure 7, la lecture de l'historique se fait en ligne. Les différentes lignes sont quant à elles ordonnées chronologiquement. Ainsi, la première ligne indique que le 8 février 2014 à 14 heures, 13 minutes, 12 secondes et 677 millisecondes (la précision temporelle dépend du système d'acquisition, mais la présence de date en millisecondes est usuelle), l'alarme identifiée « alarm12 » a disparu. Cette alarme a une criticité « LOW » et a pour description « FLTR PRESS DISCH HOPPER EAST: Level High Alarm ». Cette description doit être facilement compréhensible pour que l'opérateur comprenne tout de suite la localisation et la nature du problème.

Il est également possible de remarquer sur l'exemple que l'alarme identifiée « alarm8 » est générée (« G ») puis disparaît (« R ») sans qu'il n'y ait eu d'acquittement de la part d'un opérateur. Ceci signifie que l'opérateur n'a pas eu le temps de traiter l'alarme et que c'est le processus industriel qui, en évoluant, est sorti de son comportement anormal (par son évolution naturelle, grâce à une régulation automatique, ou grâce à des actions correctives réalisées précedemment sur d'autres alarmes), a rendu l'alarme obsolète et l'a fait disparaître. Si ce comportement est systématique, c'est qu'il s'agit typiquement d'une alarme mal configurée ou d'une information qui n'aurait pas dû être configurée en tant qu'alarme.

Il est important de noter que les informations stockées dans la base de données varient en fonction des systèmes de gestion d'alarmes car les données jugées utiles à sauvegarder sont définies dans la philosophie du système.

c. Problématique industrielle de la thèse

La modernisation de l'industrie a permis d'avoir accès plus facilement à plus d'informations sur l'évolution d'un processus industriel. Il est donc naturel que le nombre

d'alarmes configurées ait augmenté. Cependant, cette évolution a eu l'effet pervers d'augmenter la charge de travail des opérateurs qui n'ont souvent plus le temps d'analyser et de comprendre correctement les alarmes qu'ils reçoivent.

Pour pallier ce phénomène, la norme (International Society of Automation 2009) définit un cycle de vie que doit suivre tout bon système de gestion d'alarmes. Cependant, ce cycle de vie représente une démarche itérative longue qui repose en partie sur la capacité des opérateurs et des experts à comprendre facilement les alarmes.

Une première approche pour faciliter la compréhension des alarmes a été d'améliorer l'interface entre l'Homme et la machine pour que les opérateurs aient les bonnes informations au bon moment et que les experts aient une meilleure visibilité de l'évolution du processus industriel lorsqu'ils cherchent à rationaliser les alarmes. Bien qu'elle soit efficace, cette approche n'est pas suffisante car les opérateurs souffrent encore des effets d'un trop grand nombre d'alarmes à gérer rapidement. Elle n'est également pas suffisante pour identifier les nombreuses redondances entre alarmes qui sont dues à des erreurs lors de la configuration des alarmes. Il faut donc cibler ces redondances et les empêcher d'atteindre l'opérateur, sans que ce dernier ne perde d'information critique sur l'évolution du processus industriel associé. Les alarmes à éliminer seraient donc les alarmes redondantes, qui n'apportent pas d'informations supplémentaires pertinentes à l'opérateur, mais qui le sollicitent quand même.

Une seconde approche basée sur le besoin d'éliminer les alarmes redondantes a donc été proposée pour réduire à la fois le nombre et le flux d'alarmes que les opérateurs doivent traiter. Elle consiste à identifier les redondances et les liens qui unissent les alarmes avant de les analyser et de conclure sur l'utilité de ces alarmes. Pour ce faire, il faut utiliser les informations à disposition qui reflètent le travail de l'opérateur, l'évolution des alarmes ainsi que l'évolution du processus industriel. Ces informations sont celles contenues dans l'historique d'alarmes qui a été créé durant la phase de surveillance et d'évaluation du cycle de vie du système d'alarmes, ainsi que les connaissances propres du système industriel (un modèle du système par exemple).

La problématique industrielle peut donc être énoncée ainsi « Comment identifier, à partir d'un historique d'alarmes et/ou des connaissances du processus industriel, les alarmes transportant des informations redondantes pour l'opérateur qu'il est possible de supprimer sans perdre en sûreté ? ».

Ce problème fait l'objet de recherches et de communications scientifiques qui sont analysées dans le chapitre suivant. C'est également ce problème qui va être traité dans cette thèse.

4) Conclusion

Dans ce chapitre, des définitions basiques concernant la gestion d'alarmes industrielles ont été données. Le cycle de vie d'un système d'alarmes et la manière dont sont configurées et traitées les alarmes ont été décrits. Le contexte global ainsi que les problèmes liés à la mauvaise gestion des alarmes ont également été présentés.

Nous avons montré que le problème de la surabondance d'alarmes est encore d'actualité car l'amélioration continue des interfaces Homme/machine, n'est pas suffisante et présente des lacunes. Une solution qui peut contribuer à résoudre ce problème est la recherche d'alarmes qui contiennent des informations redondantes transmises à l'opérateur en vue de leur suppression. Cette suppression ne peut être effectuée que si elle ne provoque pas une perte de sûreté pour le système étudié.

Pour identifier les relations entre les alarmes, il est possible d'utiliser les données contenues dans un historique et/ou la connaissance experte du système.

Une étude des différentes publications qui traitent de ce problème ainsi que la définition de l'approche que nous proposons pour répondre à cette problématique sont détaillés dans le chapitre suivant.

Chapitre 2 : Etude bibliographique et positionnement des travaux

Dans ce chapitre, nous présentons un état de l'art des travaux relatifs à la manière d'identifier des alarmes liées par des relations physiques ou logicielles. Ces relations sont majoritairement dues à des erreurs de configuration et indiquent potentiellement l'existence de redondances entre les alarmes.

Nous nous positionnons ensuite dans cet état de l'art en indiquant les choix qui ont conduit à l'approche développée dans cette thèse pour gérer le problème du trop grand nombre d'alarmes devant être traitées par les opérateurs. Cette nouvelle approche est alors définie et détaillée. Enfin, nous reformulons ce problème pour poser les bases formelles des solutions développées dans la suite du mémoire.

1) Etat de l'Art

Les problèmes liés à la gestion des alarmes sont un centre d'intérêt pour la recherche et de nombreux papiers discutent de ce sujet. Une bonne synthèse est proposée dans (Wang et al. 2016). Ce papier fait un état de l'art et regroupe les différentes interrogations auxquelles il est nécessaire de répondre pour avoir un système d'alarmes bien géré.

Dans cet objectif, (Wang & Chen 2014) s'intéresse à la manière de gérer les alarmes bruyantes (alarmes apparaissant plusieurs fois dans un cours laps de temps) en détaillant une méthode basée sur une analyse temporelle pour les détecter et les supprimer du flux d'alarmes que doit gérer l'opérateur.

De son côté, (Yang et al. 2010) propose une manière de configurer correctement les alarmes en effectuant une analyse de corrélation entre les données historiques et les seuils de déclenchement des alarmes pour affiner et optimiser ce seuil de déclenchement.

Enfin, dans (Xu et al. 2014), les auteurs cherchent s'il est possible de prévoir les futures alarmes afin de laisser plus de temps à l'opérateur pour les traiter, en développant des modèles hybrides qui vont permettre de prédire les alarmes.

La question à laquelle nous cherchons à répondre dans cette thèse étant : « Est-il possible de réduire le nombre d'alarmes traitées par l'opérateur sans perdre en sûreté ? », nous recherchons des solutions visant à réduire le flux d'alarmes plutôt que d'augmenter le temps qu'ont les opérateurs pour traiter les alarmes. La réduction du flux d'alarmes peut se faire en recherchant des redondances et des corrélations entre occurrences d'alarmes, puis en les supprimant. Cependant, pour qualifier la perte de sûreté et pouvoir proposer des moyens pour l'empêcher, nous devons d'abord préciser ce que nous appelons sûreté.

La notion de sûreté que nous utilisons est celle définie dans la norme (Bell 1999) dans le cadre de la sûreté de fonctionnement. Elle est définie comme l'absence de risque inacceptable. Nous utilisons cette définition dans le cadre de l'analyse d'un système dans sa globalité , non limitée à l'analyse de systèmes critiques spécifiques. Ainsi, en suivant cette définition, ne pas perdre en sûreté revient à ne pas faire courir de nouveaux risques non tolérables au système. La notion de risque tolérable étant laissé à l'appréciation d'un expert.

Pour conserver la sûreté lorsque nous supprimons des alarmes, nous nous sommes inspirés de travaux adressant le problème du respect de la sûreté pour des systèmes complexes. Ainsi, (Pétin et al. 2010) montre que l'utilisation de méthodes formelles est nécessaire pour assurer la sûreté de systèmes critiques, mais que les méthodes non-formelles telles que celles utilisant le langage SysML permettent de mieux comprendre les besoins de systèmes complexes. C'est pourquoi (Pétin et al. 2010) propose une approche qui combine ces deux méthodes pour assurer la sûreté de fonctionnement d'un cas d'étude.

Dans le cas des systèmes d'alarmes, l'utilisation des historiques d'alarmes permet de construire des analyses formalisées. Ces analyses peuvent ainsi permettre de rechercher des relations entre les occurrences d'alarmes et d'assurer la conservation de la sûreté du système jusqu'au choix des alarmes à supprimer.

Nous proposons donc un état de l'art en deux parties. La première partie présente les méthodes qui n'utilisent pas d'historiques d'alarmes. Tandis que la seconde présente les méthodes qui se basent sur l'utilisation de ces historiques (seuls ou avec des données additionnelles).

a. Etudes n'utilisant pas d'historiques d'alarmes

Les premiers travaux dans ce domaine datent des années 90 ((McDonald et al. 1992) et (Wen & Chang 1997)) et présentent des algorithmes qui utilisent la connaissance du système pour relier les alarmes à des évènements anormaux. Le fait que ces recherches utilisent une association entre des ensemble d'alarmes et des perturbations leur a permis de développer des algorithmes puissants utilisant les techniques d'optimisation combinatoire. Cependant, faute d'optimisation paramétrique, ces algorithmes ont besoin d'être paramétrés manuellement. Un autre problème de ces algorithmes consiste en leur complexité incompatible avec une exploitation en ligne (surtout pour (McDonald et al. 1992) qui est cinq fois plus lent que (Wen & Chang 1997)).

Toujours utilisant la connaissance du système, (Wei et al. 2011) réalise une analyse basée sur l'architecture du système pour identifier en temps réel les fausses alarmes ainsi que les causes des perturbations et de ces fausses alarmes. Cette méthode présente l'avantage de prendre en compte différents phénomènes, mais ces travaux sont centrés autour de la norme IEC 61850 (Mackiewicz 2006) qui porte sur les protocoles de communications électroniques. La gestion des alarmes n'étant pas le cœur de ces travaux, celles-ci sont seulement catégorisées et aucune solution pour la réduction du flux d'alarmes n'est proposée.

De la même manière, (Guo et al. 2010) réalise une analyse basée sur l'architecture système avec une prise en compte de contraintes temporelles pour associer les alarmes à des défauts de fonctionnement. Cette approche prend en compte la part d'aléatoire dans la temporalité des occurrences et permet d'identifier un dysfonctionnement du système d'alarmes en cas d'alarmes manquantes. Cependant, aucune conclusion n'est tirée quant à la réduction du flux d'alarmes possible si plusieurs alarmes sont associées à la même faute.

Plutôt que de modéliser le système, (Liu et al. 2010) a préféré modéliser le comportement de l'opérateur pour évaluer les systèmes d'alarmes et évaluer leurs faiblesses. Le modèle utilisé est complet et très détaillé, mais la méthode a été testée sur un simulateur et non sur un système d'alarmes réel. Un simulateur ne représentant pas complètement la part d'incertitude et d'aléatoire, il n'est donc pas possible de conclure sur l'efficacité de cette méthode dans un cas non-simulé.

En parallèle de ces approches qui utilisent des modélisations du système et des opérateurs, des méthodes utilisant l'intelligence artificielle ont été développées.

Ainsi, (Cauvin et al. 1998) présente différentes méthodes utilisant l'intelligence artificielle au sens large de l'apprentissage automatique. Parmi celles-ci, il est possible de retrouver l'utilisation de graphes de causalité, de règles expertes, ou encore la reconnaissance de motifs. Ces méthodes sont comparées et classifiées en fonction du but recherché et de la nature du système (continu ou discret). Bien que très précise et complète, cette étude ne va pas jusqu'à proposer une méthode générique ou une généralisation des méthodes spécifiques.

Une autre étude basée sur l'utilisation de réseaux de neurones est présentée dans (De Souza et al. 2004). Dans cette étude, des réseaux de neurones entraînés sur des relations de logique floue entre les alarmes sont utilisés pour analyser des motifs de fautes et d'alarmes. L'avantage de cette approche est qu'il est possible d'analyser des motifs complexes. Cependant, celle-ci souffre des problèmes inhérents à l'utilisation de réseaux de neurones tels

que le manque de visibilité sur l'analyse réalisée par le réseau de neurones et le besoin d'avoir des périodes d'entraînement assez longues avant d'obtenir un réseau de neurones fonctionnel.

Globalement, quelle que soit l'approche utilisée, il est difficile d'avoir un modèle de connaissance complet d'un système complexe tel qu'un système d'alarmes. Cette difficulté est aggravée par le fait que se baser sur des connaissances systémiques ou s'entraîner sur un modèle spécifique empêche la généralisation des résultats.

De plus, l'élaboration de modèles physiques est techniquement difficile et prend beaucoup de temps. Par conséquent, une telle approche n'est réalisable et utile que pour un nombre limité de procédés industriels, de dispositifs ou d'équipements critiques, et n'est pas applicable de manière générique. Cette solution n'est donc pas viable lorsque l'objectif est d'obtenir une solution utilisable dans différents domaines industriels.

b. Etudes basées sur l'analyse d'historique d'alarmes

Pour développer une approche qui puisse être générique, il faut donc baser l'analyse sur autre chose que des connaissances expertes. L'utilisation d'historiques d'alarmes est une solution qui permet de développer ce type d'approche. Celle-ci peut également être complémentée par une expertise systémique.

1. Utilisation d'historiques d'alarme seulement

Certaines approches utilisent des historiques d'alarmes pour retrouver des liens causaux entre les alarmes. Par exemple, (Folmer & Vogel-Hauser 2012) regroupe des séquences apparaissant fréquemment pour identifier ces liens causaux et (Dorgo & Abonyi 2018) va plus loin en rajoutant la prise en compte de la durée des alarmes. Ces méthodes permettent d'identifier des relations simples entre des alarmes, mais ne sont pas adaptées pour trouver des relations plus complexes comme dans le cas où les alarmes ont plusieurs causes possibles ou si l'ordre d'apparition des alarmes n'est pas fixe.

Pour pouvoir identifier des relations plus complexes, (Folmer et al. 2014) considère que pour que des liens causaux existent, les alarmes doivent toujours apparaître dans le même intervalle de temps relatif. Cette approche permet d'identifier des relations temporelles en plus des relations causales. Cependant, il s'agit d'une approche statistique dont les résultats dépendent grandement du choix des paramètres. Malgré cet inconvénient, (Kinghorst et al. 2018) utilise cette approche pour rassembler graphiquement les alarmes en différents groupes appelés amas. Plus la probabilité que deux alarmes soient liées est élevée, plus ces alarmes

seront proches graphiquement. Les alarmes représentées sous forme de points et les amas qu'elles forment sont représentés sur la Figure 8 :

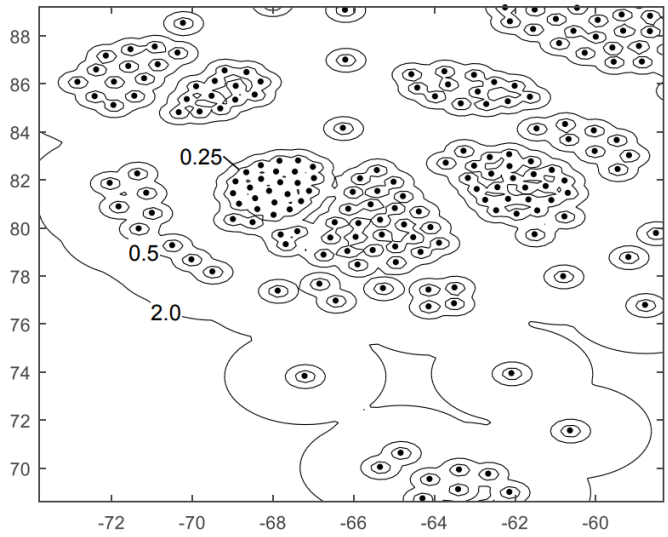


Figure 8 : Graphe de densité de probabilités que les alarmes soient liées (issu de (Kinghorst et al. 2018))

La méthode développée dans (Kinghorst et al. 2018) permet donc de regrouper des alarmes liées par des relations causales et/ou temporelles en vue d'études complémentaires ciblées sur les amas, mais elle ne permet pas de conclure par elle-même sur la réduction possible du flux d'alarmes. De plus, cette approche reste probabiliste et le choix du seuil d'isolement des amas est arbitraire. De ce fait, une incertitude plane donc sur la validité des regroupements réalisés dans les différents amas.

Plutôt que de considérer l'historique d'alarmes dans son ensemble, d'autres approches consistent à se concentrer sur les avalanches d'alarmes, étant donné que c'est durant ces avalanches que les opérateurs sont le plus débordés. Ainsi, (Cheng et al. 2013) compare deux à deux les différentes avalanches d'alarmes pour trouver des séquences d'alarmes communes entre elles. Cette approche prend en compte l'analyse du temps lorsque des alarmes apparaissent proches l'une de l'autre, et elle permet d'identifier les causes d'une avalanche d'alarmes et de réduire le flux d'alarmes lorsque celles-ci se produisent. Cependant, cette méthode ne permet pas d'identifier des relations entre alarmes isolées et ne permet donc pas de conclure quant à une réduction du flux d'alarmes sur l'historique global.

Toujours en se concentrant sur les avalanches d'alarmes, (Ahmed et al. 2013) recherche des similarités entre elles pour les catégoriser. (Fullen et al. 2017) va plus loin en utilisant différentes notions de distance pour définir formellement ces similarités. Ces deux approches permettent d'identifier des phénomènes complexes tels que les effets de perturbations lors

d'avalanches d'alarmes, mais l'absence de conclusion sur les alarmes isolées rend plus difficile l'obtention des conclusions expertes nécessaires pour quantifier la réduction du flux d'alarmes.

Une autre vision consiste à considérer les alarmes comme des variables binaires égales à 1 quand l'alarme est active et égale à 0 quand l'alarme est inactive. C'est le cas de (Kondaveeti et al. 2012) qui cartographie les similarités entre ces variables d'alarme binaires. Un exemple de cartographie obtenue par cette méthode est donné sur la Figure 9 :

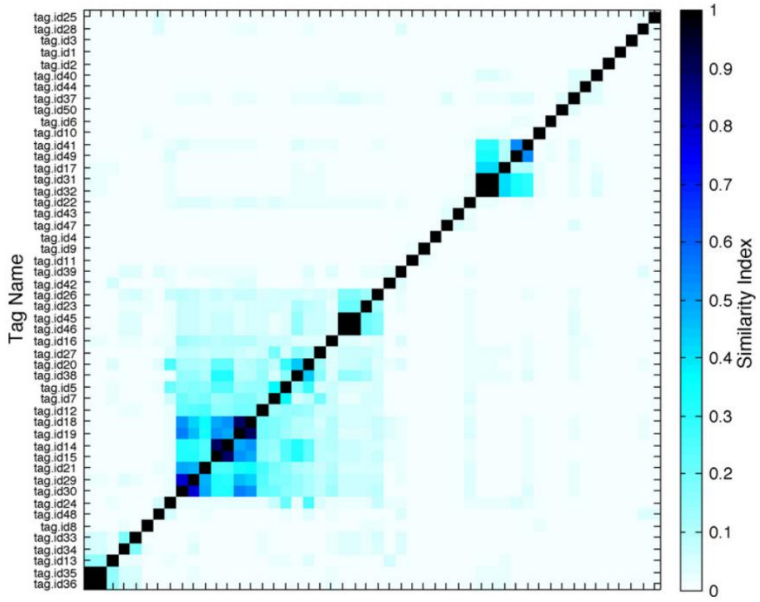


Figure 9 : Cartographie de la similarité entre alarmes (issu de (Kondaveeti et al. 2012))

Sur cette figure, les identifiants des alarmes sont indiqués à gauche et servent en abscisse et en ordonnée. L'index coloré du taux de similarité est indiqué à droite. En première analyse, la diagonale foncée représente le fait qu'une alarme à un taux de similarité de 1 avec elle-même. Cette cartographie permet de visualiser facilement les alarmes redondantes ou en relations en indiquant un niveau de certitude sur l'existence des relations (plus les carrés sont foncés, plus l'existence de la relation est probable).

De plus, l'approche proposée par (Kondaveeti et al. 2012) prend en compte le temps en considérant la durée des alarmes, leur date d'apparition et leur date de disparition. Cependant, cette étude se limite à des scénarios où des comportements anormaux tels que des avalanches d'alarmes se produisent, et elle ne permet pas d'identifier la nature des relations entre alarmes.

Poursuivant les travaux de (Kondaveeti et al. 2012), (Lucke et al. 2018) utilise les similarités identifiées pour prédire de futurs comportements anormaux. En plus des considérations temporelles, cette approche prend en compte des phénomènes supplémentaires tels que la

présence d'alarmes bruyantes. Ces travaux permettent d'analyser des alarmes avant qu'elles n'atteignent l'opérateur, mais pour cela le modèle a besoin d'être entraîné sur un nombre de données suffisant. De plus, comme pour la méthode précédente, aucune information concernant la nature des relations n'est disponible.

De son côté, (Yang et al. 2012) adapte l'approche de (Kondaveeti et al. 2012) pour effectuer une analyse sur l'ensemble de l'historique d'alarmes et regrouper les alarmes au sein d'une cartographie colorée selon des coefficients de corrélation. Cependant, comme pour les travaux précédents, aucune information sur la nature des corrélations identifiées n'est disponible.

Plutôt que de représenter graphiquement les similarités, (Yang et al. 2013) préfère les associer à la notion de retards de corrélation (le fait que les alarmes ont un comportement dynamique et ne se comportent pas toujours de la même manière) pour prendre en compte le temps et ainsi identifier des corrélations plus complexes. Cette méthode permet également de mieux résister au bruit créé par des problèmes d'acquisition que la méthode de (Kondaveeti et al. 2012). Cependant, il n'est toujours pas possible d'obtenir des informations sur la nature des relations entre alarmes en utilisant cette approche.

Toujours en utilisant la notion de retard de corrélation, (Wang et al. 2015) cherche les liens causaux entre les alarmes pour retrouver l'origine des alarmes. Ces travaux permettent d'identifier des relations complexes en prenant en compte le temps, mais les résultats manquent de certitude étant donné qu'une approche probabiliste par distribution est utilisée pour déterminer les relations de causalité.

Ainsi que nous l'avons constaté, les approches basées sur l'utilisation de données historiques permettent d'identifier des relations complexes entre alarmes. Cependant, elles sont souvent lourdes en temps de calcul et ne permettent que rarement de comprendre les phénomènes responsables de ces relations. De plus, pour valider les résultats des analyses et pour conclure sur une réduction du flux d'alarmes qui permet de conserver la sûreté du système, il est très fréquemment nécessaire d'avoir des connaissances sur le processus industriel associé au système d'alarmes.

2. Utilisation de données additionnelles à l'historique d'alarmes

Pour pouvoir comprendre et valider les relations identifiées par différentes approches, il est souvent nécessaire d'incorporer une part de connaissance experte. Ainsi, (Dorgo et al. 2018) met en commun les informations contenues dans la base de données des actions opérateurs avec l'historique d'alarmes pour effectuer une analyse heuristique. Cette approche permet de trouver les relations qui existent entre les alarmes et les actions effectuées par l'opérateur pour

répondre à ces dernières. Avec cette méthode, il est donc possible d'identifier les raisons ayant mené les opérateurs à prendre des actions précises et les effets qu'ont eu les réponses des opérateurs sur le système. Cette étude est cependant longue et la partie qualitative de l'analyse due à l'interaction avec des opérateurs limite la généralisation de ses résultats.

Plutôt que d'augmenter le volume de données à analyser, (El Sakka et al. 2018) fait le choix d'utiliser des connaissances du système en prétraitement avant l'analyse des données historiques. L'utilisation de connaissances expertes telles que les associations entre les alarmes et les équipements permet de définir des groupements d'alarmes qui contiennent les alarmes liées par des relations causales. Cette approche permet ainsi d'éviter d'identifier de fausses relations causales. Cependant, cette étude ne présente pas de cas d'application de cette méthode et il peut être intéressant de généraliser l'utilisation d'un prétraitement à la recherche d'autres phénomènes que des relations causales.

Un autre type de prétraitement est proposé dans (Vasquez Capacho et al. 2017) : Une expertise est utilisée pour construire les séquences d'alarmes à analyser à partir de différents scénarios. Cette approche prend en compte le temps et permet de modéliser des comportements pouvant déclencher un grand nombre d'alarmes. Le cas d'application utilisé dans cette étude est une raffinerie pétrochimique colombienne et montre que l'algorithme utilisé passe correctement à l'échelle. Malgré ses bons résultats, cette approche peut être encore améliorée en prenant en compte d'autres informations systémiques pour obtenir une modélisation plus complète du comportement des alarmes. L'inconvénient de celle-ci est qu'il est difficile de construire des séquences complètes qui représentent correctement l'évolution du processus.

De leur côté, (Abele et al. 2013) et (Wang et al. 2014) utilisent des réseaux bayésiens dont la structure est connue pour représenter l'historique d'alarmes et identifier l'origine des alarmes. Par la suite, les résultats obtenus par cette approche sont comparés dans (Abele et al. 2013) à des résultats obtenus par une approche par apprentissage. Cette approche a l'avantage de prendre en compte une incertitude sur la connaissance experte. Cependant, la construction des réseaux bayésiens est longue et complexe. Ce problème a d'ailleurs conduit à comparer l'efficacité des différentes manières de construire ce type de réseaux dans (Wunderlich & Niggemann 2017).

De son côté, (Simeu-Abazi et al. 2011) préfère utiliser des arbres de fautes pour lier les alarmes et les fautes qui les précèdent. Cette approche permet de filtrer les redondances d'alarmes ainsi que les fausses alarmes tout en prenant en compte la dynamique du système.

Cependant, malgré une démonstration de cette méthode sur un cas réel, des moyens de modélisation insuffisants rendent les conclusions vis-à-vis du système réel incomplètes.

Toujours dans l'idée d'utiliser la connaissance des fautes pour identifier des liens entre les alarmes, (Schleburg et al. 2013) utilise la connaissance du système pour formuler des règles qui sont confrontées à l'historique d'alarmes. Grâce à cette approche, il est possible de retrouver l'origine du problème et de suivre la propagation des anomalies ou des fautes dans le système. En regroupant les alarmes reliées qui correspondent à un même problème, la charge de travail des opérateurs peut être réduite. Cette technique présente cependant des limitations dues au fait qu'elle nécessite de nombreuses itérations et que la construction des règles ainsi que leur syntaxe sont spécifiques à chaque système.

Le suivi de la propagation des fautes ou des anomalies dans un système permet de comprendre les relations causales entre alarmes. Dans cette optique, (Chiang & Braatz 2003) allie une analyse statistique pour créer une carte de causalité entre toutes les variables mesurées du système. Cette approche prend en compte la dynamique du système pour obtenir une meilleure détection des fautes et un meilleur suivi de leur propagation. Cependant, le fait d'appliquer une étude statistique réalisée sur un simulateur ne prend pas en compte tous les phénomènes possibles d'un système réel. De ce fait, cette étude ne permet pas de valider ses résultats pour un système réel et d'extrapoler les conclusions sur le comportement des alarmes.

Plutôt que de mettre en parallèle l'historique d'alarmes avec des phénomènes du système tels que des fautes, (Thornhill et al. 2003) met en avant l'utilisation des connaissances du processus possédées par des experts pour améliorer et valider les résultats de l'analyse automatique de l'historique. Cette méthode permet une meilleure identification des causes des alarmes et aide à la prise de décision des opérateurs, mais se limite au seul cas de la propagation d'un problème dans le système.

L'approche par la seule analyse d'historiques d'alarmes permet donc d'effectuer des études génériques qui peuvent être transposées sur différents systèmes et qui identifient potentiellement des phénomènes complexes. Cette approche permet de construire des méthodes formalisées nécessaires pour assurer la conservation de la sûreté du système. Cependant, l'utilisation de connaissances expertes est également indispensable car elle permet de considérer le système dans sa globalité et de valider les résultats de ces méthodes. Elle permet donc de conclure sur la réduction effective du flux d'alarmes tout en s'assurant de conserver la sureté du système.

2) Approche proposée

a. Méthode générale

L'état de l'art a permis de comparer les différentes approches contribuant à la réduction du flux d'alarmes, ainsi qu'à mettre en évidence leurs possibilités et limites. Notre objectif étant de développer une approche générique, nous avons fait le choix d'exclure l'utilisation des informations qui dépendraient des opérateurs et des outils à leur disposition. Notre approche est donc basée uniquement sur l'utilisation d'historiques d'alarmes sans autres informations additionnelles pour identifier les relations entre les alarmes contenues dans cet historique. Ces relations sont ensuite confiées à un expert capable de les valider et de conclure sur la réduction du flux d'alarmes possible sans perte de sûreté.

Toujours dans l'objectif de préserver la sûreté du système, nous avons choisi de ne pas réaliser d'étude probabiliste et de tester notre approche sur des données réelles provenant de systèmes physiques plutôt que d'utiliser des données provenant de simulateurs. En effet, les simulateurs ne prennent pas en compte des comportements imprévus par leurs concepteurs et ne sont donc pas suffisants pour valider l'efficacité de la méthode pour un système réel.

La structure générale de l'approche proposée et la manière dont celle-ci s'intègre dans un système de gestion d'alarmes sont représentés sur la Figure 10 :

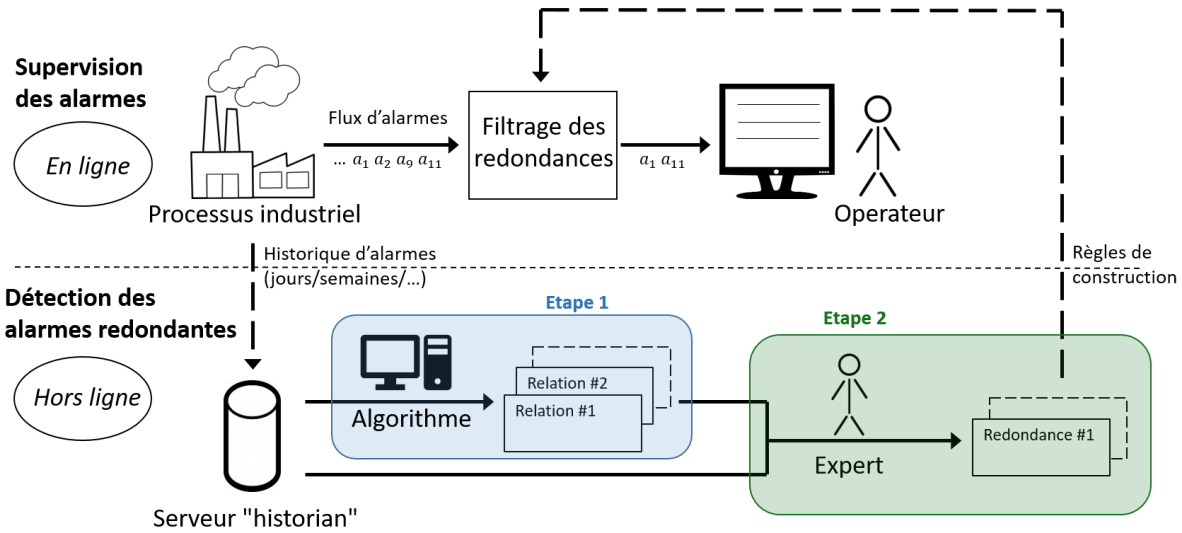


Figure 10 : Schéma détaillant l'approche proposée

Notre approche s'intègre dans un système d'alarme classique où les alarmes générées par un processus industriel sont directement transmises à un opérateur et où elles sont régulièrement stockées dans une base de données historique. L'idée derrière notre approche

est de réduire le flux d'alarmes arrivant aux opérateurs en filtrant en temps réel les alarmes redondantes. Pour ce faire, le filtre doit posséder des règles prédéfinies contenant la connaissance des redondances dont la suppression n'entraîne pas une perte de sûreté pour le processus industriel.

Ces redondances peuvent être obtenues grâce à une analyse de l'historique d'alarmes. Nous proposons d'effectuer cette analyse en deux étapes consécutives :

- La première étape utilise des méthodes formalisées pour identifier automatiquement des relations entre les alarmes, et ainsi déterminer des alarmes redondantes qui pourraient être supprimées sans perte d'information sur la logique de leurs occurrences ;
- La seconde étape consiste à confier ces relations et ces propositions de suppression à un expert pour qu'il puisse choisir les alarmes redondantes qui vont être supprimées sans provoquer de perte de sûreté. Durant son analyse, l'expert garde une vision globale du comportement des alarmes en considérant l'historique d'alarmes dans son ensemble.

Nos travaux se concentrent sur la première étape et consistent à identifier des relations entre alarmes et proposer des suppressions à un expert. Nous ne traitons ni du détail de l'expertise servant à identifier les alarmes dont la suppression n'entraîne pas de perte de sûreté, ni de la mise en œuvre du filtrage sur le flux d'alarmes.

b. Formulation grâce aux paradigmes des Systèmes à Evènements Discrets

Pour identifier ces relations, nous nous inspirons des techniques de recherche de relations entre évènements qui existent dans le domaine des Systèmes à Evènements Discrets (SED). Il est possible de se rapprocher de ce domaine en modélisant les alarmes comme des évènements. En utilisant cette représentation, le système industriel peut être considéré comme un générateur d'évènements.

Les alarmes étudiées sont issues d'un historique contenant des informations telles que les dates d'apparition et de disparition des alarmes. En modélisant l'apparition d'une alarme comme un évènement et la disparition d'une alarme comme un autre évènement, il est donc possible d'associer à chacun de ces évènements une date d'occurrence.

Sur l'exemple de configuration d'alarme donné sur la Figure 11, il est possible de voir que l'alarme associée à une mesure capteur est apparue deux fois et a disparu une fois. Ainsi,

l'évènement « apparition de l'alarme » possède deux occurrences et l'évènement « disparition de l'alarme » en possède une seule.

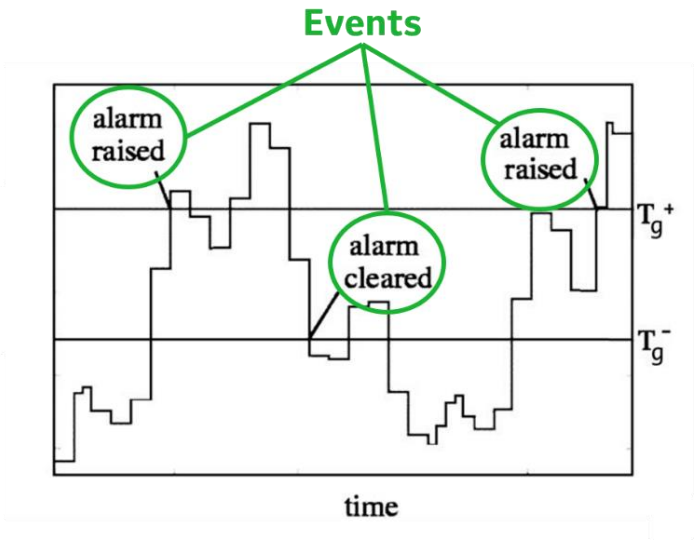


Figure 11 : Exemple de configuration d'alarme

Etant donné que la disparition des alarmes peut dépendre de la réponse d'un opérateur, seules les apparitions d'alarmes peuvent être considérées comme uniquement dépendantes du système et indépendantes de toute action humaine. C'est pourquoi nous n'utilisons dans la suite que les seules apparitions d'alarmes pour notre étude.

De plus, toujours pour proposer une méthode générique qui ne dépende pas du système étudié, le choix a été fait de ne pas prendre en compte des informations autres que l'identifiant des alarmes, telles que le niveau de criticité des alarmes, qui peuvent se trouver dans l'historique. En effet, ce type d'information dépend de la philosophie du système d'alarmes et est donc spécifique à chaque système.

Nous ne prenons également pas en compte la dates d'occurrences des différents évènements. Ce choix de ne pas prendre en compte le temps physique est motivé par le fait que, dans un premier temps, nous explorons la faisabilité et l'efficacité de notre approche dans un contexte industriel.

Pour alléger la suite du texte, et étant donné que nous réduisons l'étude de l'historique d'alarmes à l'étude des apparitions d'alarmes, par abus de langage, nous parlerons d'"alarmes" pour qualifier l'"apparition d'une alarme" et son évènement correspondant.

Le problème de la réduction du flux d'alarmes revient donc à diminuer le nombre d'occurrences d'alarmes qui atteignent l'opérateur. Nous cherchons donc à réduire le nombre

d'occurrences d'évènements en recherchant des relations entre ces évènements à partir d'un historique constitué d'une séquence d'évènements. Pour ce faire, les hypothèses qui sont sous-jacentes à notre approche sont la non simultanéité des occurrences d'évènements, l'absence de connaissance a priori sur le processus industriel, et la non prise en compte du temps physique.

1. *Définitions*

Pour poser les bases de la méthode développée pour répondre à ce problème, il faut définir certains termes que nous utilisons :

Définition 1 : *Séquence d'alarmes* :

Une séquence d'alarmes générée par un processus industriel est une succession ordonnée d'occurrences d'alarmes.

La longueur d'une séquence d'alarmes S est notée $|S|$ et est le nombre d'occurrences que contient S. Par exemple, la séquence donnée sur la Figure 12 est de longueur 14 et peut être écrite $S = a_1a_5a_2a_5a_1a_8a_2a_8a_1a_2a_5a_5a_1a_2$.

Définition 2 : *Alphabet d'alarmes* :

L'ensemble de toutes les alarmes a_i dont des occurrences composent la séquence S est appelé l'alphabet d'alarmes (ou simplement alphabet) de S et est noté Σ_S. De plus, Σ_S^* est l'ensemble de toutes les séquences qu'il est possible de construire sur Σ_S.

Sur la séquence donnée sur la Figure 12, l'alphabet d'alarmes est $\Sigma_S = \{a_1, a_2, a_5, a_8\}$.

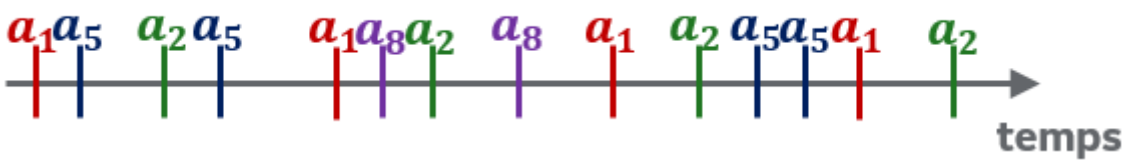


Figure 12 : Exemple de séquence d'alarmes

Pour réduire le flux d'alarmes, nous cherchons à supprimer certaines alarmes qui sont redondantes. Ces redondances vont pouvoir être déduites de relations entre alarmes que nous obtenons grâce à différentes méthodes formalisées.

L'ensemble des relations r entre occurrences d'alarmes est noté R, et Σ_r est ensemble des alarmes dont une ou plusieurs occurrences sont impliquées dans r. La forme de r varie en fonction des méthodes utilisées mais sa nature qui est l'expression de relations de précédence entre des occurrences d'alarmes reste la même.

Pour définir correctement cette notion d'équivalence entre séquences, nous avons besoin de définir quelques termes supplémentaires :

Définition 3 : *Concaténation* (Hopcroft et al. 2003):

Soit 2 séquences x et y, xy est la concaténation de x et de y. C'est-à-dire, la séquence formée en faisant suivre x par y. Plus précisément, si x est la séquence composée de i évènements tel que $x = x_1x_2 \dots x_i$ et que y est la séquence composée de j évènements tel que $y = y_1y_2 \dots y_j$, alors xy est la séquence de longueur $i + j$: $xy = x_1x_2 \dots x_iy_1y_2 \dots y_j$.

Dans ce contexte, la séquence vide d'évènement ε est l'élément neutre de la concaténation : $\varepsilon x = x\varepsilon = x$

Définition 4 : *Projection d'une séquence* :

Selon (David & Alla 2005), la projection d'une séquence S sur un alphabet Σ, notée $Proj_{\Sigma}(S)$, consiste à remplacer tous les a_i appartenant à Σ_S, mais pas à Σ, par l'élément neutre de la concaténation ε.

En appliquant cette définition sur la séquence exemple S de la Figure 12, avec un alphabet de projection $\Sigma = \{a_1, a_8\}$, nous obtenons :

$$Proj_{\Sigma}(S) = a_1\varepsilon\varepsilon\varepsilon a_1a_8\varepsilon a_8a_1\varepsilon\varepsilon\varepsilon a_1\varepsilon = a_1a_1a_8a_8a_1a_1$$

Définition 5 : *Equivalence de deux séquences* :

Deux séquences S et $\tilde{S}$ sont équivalentes relativement à un ensemble R de relations identifiées si, et seulement si, pour toutes les relations identifiées r, une partie de la projection sur Σ_r d'une des séquences est incluse dans la projection sur Σ_r de l'autre séquence. Cette équivalence est notée $(S, R) \equiv (\tilde{S}, R)$:

$$(S,R) \equiv \left(\tilde{S},R\right) <=> \forall r \in R, \exists\, \tilde{\sigma} \in \Sigma_r^+ \,/ \begin{cases} Proj_{\Sigma_r}\left(\tilde{S}\right) = \tilde{S}_{pré}\tilde{\sigma}\tilde{S}_{suff} \\ (\tilde{S}_{pré}, \tilde{S}_{suff}) \in (\Sigma_r^*)^2 \\ \exists\, \sigma \in \Sigma_r^+ \,/ \begin{cases} Proj_{\Sigma_r}(S) = S_{pré}\sigma S_{suff} \\ (S_{pré}, S_{suff}) \in (\Sigma_r^*)^2 \\ \begin{cases} \sigma = \sigma_{pré}\tilde{\sigma}\sigma_{suff} \\ (\sigma_{pré}, \sigma_{suff}) \in (\Sigma_r^*)^2 \end{cases} \\ OU \\ \begin{cases} \tilde{\sigma} = \tilde{\sigma}_{pré}\sigma\tilde{\sigma}_{suff} \\ (\tilde{\sigma}_{pré}, \tilde{\sigma}_{suff}) \in (\Sigma_r^*)^2 \end{cases} \end{cases} \end{cases}$$

Dans cette définition, les séquences projetées $Proj_{\Sigma_r}\left(\tilde{S}\right)$ et $Proj_{\Sigma_r}(S)$ sont représentées par $\tilde{S}_{pré}\tilde{\sigma}\tilde{S}_{suff}$ et $S_{pré}\sigma S_{suff}$. Cela permet de faire apparaître les séquences $\tilde{S}_{pré}$ et $S_{pré}$ qui sont

appelées préfixes des séquences $Proj_{\Sigma_r}(\tilde{S})$ et $Proj_{\Sigma_r}(S)$. De même, $\tilde{S}_{suff}$ et S_{suff} sont appelées suffixes des séquences $Proj_{\Sigma_r}(\tilde{S})$ et $Proj_{\Sigma_r}(S)$. Ces notations sont utilisées pour indiquer que $\tilde{\sigma}$ et σ doivent être des sous-séquences de $Proj_{\Sigma_r}(\tilde{S})$ et $Proj_{\Sigma_r}(S)$.

Exemple :

Pour illustrer la notion d'équivalence de deux séquences, nous considérons le cas simple de la séquence d'alarmes $S = a_1a_2a_5a_8$. Nous considérons également qu'il n'existe qu'une seule relation r entre alarmes, et que celle-ci lie les alarmes a_2 et $a_5 : R = \{r\}$ avec $\Sigma_r = \{a_2, a_5\}$.

- Dans le cas où $\tilde{S} = a_1a_2a_8$: $Proj_{\Sigma_r}(S) = a_2a_5$ et $Proj_{\Sigma_r}(\tilde{S}) = a_2$

Il existe $\tilde{\sigma} = a_2$ tel que $\sigma = a_2a_5$ puisse s'écrire de la forme $\sigma = \varepsilon\tilde{\sigma}a_5$ avec $(\varepsilon, a_5) \in (\Sigma_r{}^*)^2$ et $(\tilde{\sigma}, \sigma) \in (\Sigma_r^+)^2$. La définition de l'équivalence est donc respectée relativement à $\{r\}$ et $(S, \{r\}) \equiv (\tilde{S}, \{r\})$.

- Dans le cas où $\tilde{S} = a_1a_8$: $Proj_{\Sigma_r}(S) = a_2a_5$ et $Proj_{\Sigma_r}(\tilde{S}) = \varepsilon$

La seule sous-séquence de $Proj_{\Sigma_r}(\tilde{S})$ est $\tilde{\sigma} = \varepsilon \notin \Sigma_r^+$. Il n'existe donc pas de $\tilde{\sigma} \in \Sigma_r^+$. De ce fait, dans ce cas, S et $\tilde{S}$ ne sont pas équivalent relativement à $\{r\}$.

2. *Les méthodes formalisées utilisées pour réduire le flux d'alarmes*

Grâce à la connaissance de l'ensemble des relations entre alarmes R, nous proposons des séquences équivalentes $\tilde{S}$ qui ne perdent pas d'information concernant la logique des occurrences par rapport à une séquence d'alarmes S. Comme l'objectif final est de diminuer le flux d'alarmes, nous souhaitons remplacer S par $\tilde{S}$ tout en ayant $(S, R) \equiv (\tilde{S}, R)$ et $|\tilde{S}| < |S|$.

Cependant, la définition de l'équivalence de deux séquences fait qu'il existe une infinité de séquences $\tilde{S}$ telles que $(S, R) \equiv (\tilde{S}, R)$. C'est à ce stade que commence la seconde étape de notre approche et que l'expert va devoir choisir, grâce à sa connaissance de R et du système, quelle séquence est celle qui permet la meilleure réduction tout en conservant la sûreté du système. Cette séquence est choisie parmi l'ensemble des $\tilde{S}$ qu'il est possible d'obtenir depuis S par suppression d'occurrences (voulant réduire le nombre d'occurrences, nous ne nous intéressons qu'à des suppressions).

Le problème que nous cherchons à résoudre dans nos travaux est celui de l'identification des différentes relations existant entre les occurrences d'alarmes. Pour ce faire, deux méthodes sont utilisées dans la suite :

- Une méthode par recherche d'adjacences fréquentes : Avec cette méthode, nous identifions les occurrences d'alarmes qui sont fréquemment adjacentes ;
- Une méthode par recherche de précédences systématiques : Avec cette méthode, nous identifions des comportements systématiques où les occurrences d'une alarme précèdent toujours les occurrences d'une autre alarme sans qu'elles ne soient obligatoirement adjacentes.

L'exploration de ces deux méthodes commence par l'approche basée sur la recherche d'adjacences fréquentes car celle-ci peut se ramener à la recherche de séquences fréquentes au sein de la séquence d'alarme et que ce type de technique est fréquemment rencontré dans la littérature spécialisée dans la réduction des flux d'alarmes.

Chapitre 3 : Identification d’adjacences fréquentes entre alarmes

Dans ce chapitre, pour identifier les adjacences fréquentes (alarmes apparaissant souvent côte à côte) qui existent au sein de la séquence d'alarmes, une méthode consistant à y rechercher des séquences qui se répètent fréquemment est détaillée. Ces adjacences peuvent être étudiées indépendamment et elles permettent de proposer à un expert des séquences équivalentes permettant potentiellement de réduire le flux d'alarmes sans perdre en sûreté.

Après avoir formalisé le problème, nous explorons différents algorithmes existants qui peuvent permettre d'identifier des séquences fréquentes. Puis nous détaillons l'algorithme développé autour de l'algorithme *AprioriAll* présenté dans (Agrawal & Srikant 1995). Notre algorithme est ensuite appliqué à un exemple que nous détaillons et les résultats obtenus sont analysés pour conclure sur l'efficacité de cette méthode.

1) Réduction de la séquence par analyse de motifs fréquents

L'objectif de la méthode développée dans ce chapitre est de déterminer une séquence équivalente qui, grâce à la connaissance des adjacences fréquentes, ne perde pas la logique des occurrences contenue dans la séquence d'alarmes historique.

L'hypothèse qui a guidé le développement de cette méthode est que des alarmes redondantes se retrouvent fréquemment côte à côte dans la séquence d'alarmes et forment donc des adjacences fréquentes. Des alarmes adjacentes formant une séquence, pour identifier les adjacences fréquentes, nous cherchons les séquences qui apparaissent fréquemment dans la séquence d'alarmes.

a. Support d'un motif et motif fréquent

Pour pouvoir développer l'analyse permettant d'identifier les adjacences fréquentes, nous définissons les notions de motif, de support d'un motif et de motif fréquent:

Définition 1 : M*otif* :

Dans le cas de notre étude, un motif est une séquence d'évènements particulière.

Définition 2 : *Support d'un motif* :

Le support d'un motif est le nombre de fois que le motif apparaît dans le jeu de données considéré. Dans notre cas, c'est le nombre de fois que la séquence servant de motif est présente au sein de la séquence d'alarmes globale S et est noté $Supp_S()$.

Définition 3 : *Motif fréquent* (Han et al. 2011) :

Les motifs fréquents sont des motifs qui ont une valeur de support supérieur à un seuil propre à chaque analyse. Dans notre cas, étant donné l'absence de connaissance a priori, pour ne pas manquer de relations d'adjacence fréquente, nous avons fait le choix générique de considérer fréquent tout motif qui a un support d'au moins 2.

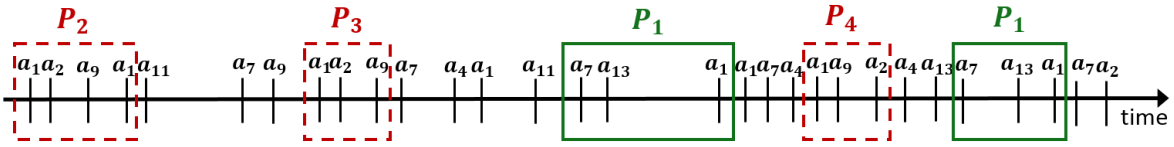


Figure 13 : Exemples de motifs

La Figure 13 est une séquence d'alarmes sur laquelle plusieurs exemples de motifs sont illustrés. La séquence motif $P_1 = a_7 a_{13} a_1$, encadrée en trait plein, est présente deux fois dans la séquence d'alarmes et a donc un support de deux. Dans le cadre de notre étude, cette valeur de support est suffisante pour qualifier P_1 de motif fréquent.

D'autres motifs sont représentés en pointillés sur la Figure 13, ces différents motifs sont $P_2 = a_1 a_2 a_9 a_1$, $P_3 = a_1 a_2 a_9$ et $P_4 = a_1 a_9 a_2$. Ces trois motifs sont effectivement différents car P_2 a une occurrence de a_1 en plus que P_3 et les occurrences de a_2 et a_9 sont alternés entre P_3 et P_4. Bien que ces trois motifs possèdent le même alphabet d'alarmes, leur séquence sont différentes et ils sont donc traités indépendamment lors du calcul de la valeur de leur support.

b. Obtention des séquences équivalentes

Les motifs fréquents représentant les relations d'adjacence fréquente, si P est l'ensemble des motifs fréquents, nous proposons à l'expert les séquences équivalentes $\tilde{S}$ telles que $(S, P) \equiv (\tilde{S}, P)$ et $|\tilde{S}| < |S|$.

Etant donné qu'une séquence S peut toujours être représentée comme la concaténation de trois séquences telle que $S = \sigma_{Pré}\sigma\sigma_{Suff}$ avec $|S| = |\sigma_{Pré}| + |\sigma| + |\sigma_{Suff}|$, la diminution de la longueur de σ revient à diminuer la longueur de S. Dans le cas de notre étude, σ se trouve être un motif fréquent. L'analyse sur les motifs fréquents permet donc de conclure sur la diminution de longueur totale de la séquence d'alarmes : la réduction de la longueur de S correspondant à la réduction réalisée sur un motif multipliée par la valeur du support de ce motif.

Il est donc possible de considérer indépendamment les différents motifs fréquents p pour obtenir des séquences équivalentes $\tilde{p}$ qui respectent la relation suivante :

$$\tilde{p} \mid \begin{cases} (p,\{p\}) \equiv (\tilde{p},\{p\}) \\ |\tilde{p}| \leq |p| \end{cases} \quad \text{avec} \quad \begin{cases} |p| \geq 2 \\ Supp_S(p) \geq 2 \end{cases}$$

Il est ensuite possible d'obtenir les séquences équivalentes $\tilde{S}$ à partir de S en remplaçant, pour tous les motifs fréquents de P, des occurrences du motif p par son motif équivalents $\tilde{p}$.

2) Algorithme d'identification des motifs fréquents

Pour fournir les relations d'adjacence fréquente, il nous faut un algorithme capable d'identifier des séquences qui se répètent au sein de la séquence d'alarmes.

a. Choix de l'algorithme

Notre méthode consistant à identifier des séquences s'inclut dans le domaine de recherche plus large qu'est la fouille de données et ne se limite pas à la seule problématique de réduction du nombre d'alarmes industrielles. Ainsi, comme l'indique (Zhao & Bhowmick 2003), la majorité des algorithmes de recherche de séquences identifient des relations entre les éléments d'une base de données, peu importe sa nature.

Les différentes approches peuvent se regrouper en deux catégories : les méthodes basées sur des approches Apriori et les méthodes basées sur la croissance de motifs.

1. Approches Apriori

Les premiers travaux ayant pour but d'identifier des séquences d'intérêt sont présentés dans (Agrawal et al. 1993). L'approche présentée est une approche ascendante où des motifs sont étendus par un élément à la fois pour générer des candidats. Ces candidats sont des potentiels motifs et leur existence dans le jeu de données est vérifiée avant de passer à l'incrément suivant. L'algorithme se termine lorsqu'il n'est plus possible de faire croître les motifs.

Cette approche a le mérite d'être simple à mettre en place mais souffre d'une complexité exponentielle due à la génération et au test des différents candidats.

2. Approches par croissance de motifs

Pour éviter les problèmes liés à la complexité algorithmique de l'approche Apriori, une autre approche qui enlève le besoin de générer des candidats a été développée.

Cette approche est présentée dans (Han et al. 2000). Elle comprime la base de données en un arbre de motifs fréquents qui est ensuite séparé en plusieurs bases de données obtenues par projection. Les analyses sur ces nouvelles bases de données sont ensuite réalisées indépendamment les unes des autres.

Cette méthode est donc beaucoup plus rapide quand il y a un grand volume de données à analyser. Cependant, cette méthode est également plus compliquée à développer, tester et maintenir.

3. *Algorithme sélectionné*

Etant donné que dans notre méthode la recherche de motifs fréquents est effectuée hors-ligne, le temps de calcul n'est pas un critère bloquant. De plus, des considérations industrielles de mise en place et d'utilisation nécessitent l'utilisation d'un algorithme relativement simple. C'est pourquoi l'algorithme qui a été retenu est celui appelé Apriori.

Plus précisément, comme nous souhaitons identifier toutes les relations possibles, l'algorithme développé est basé sur l'algorithme *AprioriAll* présenté dans (Agrawal & Srikant 1995).

b. Description de *AprioriAll*

L'algorithme *AprioriAll* a été développé pour analyser des paniers de consommateurs. Dans (Agrawal & Srikant 1995), l'objectif est d'identifier des suites de produits qui sont souvent achetés les uns à la suite des autres. Ainsi, dans ce cadre, les motifs fréquents sont des suites fréquentes d'achats. Ces suites sont représentées sous la forme de séquences. Ainsi, si un client achète le produit A et que plus tard il achète le produit B, cette suite d'achats est notée AB.

Le fonctionnement de *AprioriAll* est basé sur les deux axiomes suivants :

- Toutes les séquences extraites d'une séquence fréquente sont fréquentes ;
- Si une séquence n'est pas fréquente, les séquences dont elle est préfixe ne sont pas fréquentes.

Outre l'étape de formatage des données, l'algorithme peut être découpé en quatre phases différentes.

Tableau 2 : Produits achetés par transaction et par client (inspiré de (Agrawal & Srikant 1995))

	Produits achetés par transaction et par client			
	Transaction 1	Transaction 2	Transaction 3	Transaction 4
Client 1	{A, E}	{B}	{C}	{D}
Client 2	{A}	{C}	{D}	{C, E}
Client 3	{A}	{B}	{C}	{D}
Client 4	{A}	{C}	{E}	
Client 5	{D}	{E}		

Le Tableau 2 montre des données issues de (Agrawal & Srikant 1995) qui vont servir à illustrer les différentes phases de l'algorithme. Les achats de cinq produits A, B, C, D et E sont classés par transaction et par client.

1. *Phase d'identification des séquences élémentaires*

Lors de cette phase, les données sont analysées pour identifier l'alphabet des différents produits qui sont achetés fréquemment et pour construire les séquences élémentaires (de longueur 1) composées de ces produits.

Le Tableau 3 montre les séquences élémentaires obtenues à partir des données présentées dans le Tableau 2 :

Tableau 3: Séquences élémentaires (inspiré de (Agrawal & Srikant 1995))

Séquences élémentaires	Support
A	4
B	2
C	4
D	4
E	4

La valeur du support correspond au nombre de suites d'achats dans lesquelles les produits de ces séquences ont été trouvés. C'est-à-dire le nombre de clients qui ont acheté ces produits.

2. *Phase de transformation en suite d'achats*

L'objectif de cette phase est de traduire les suites d'achats sous forme de séquences. Le Tableau 4 montre les résultats de cette phase appliquée au cas des données du Tableau 2 :

Tableau 4 : Suites d'achats (inspiré de (Agrawal & Srikant 1995))

	Produits achetés par transaction et par client				**Suites d'achats**
	Transaction 1	Transaction 2	Transaction 3	Transaction 4	
Client 1	{A, E}	{B}	{C}	{D}	**{A, E} B C D**
Client 2	{A}	{C}	{D}	{C, E}	**A C D {C, E}**
Client 3	{A}	{B}	{C}	{D}	**A B C D**
Client 4	{A}	{C}	{E}		**A C E**
Client 5	{D}	{E}			**D E**

Ces suites d'achats sont ensuite représentées comme des séquences où les produits achetés lors d'une même transaction client sont regroupés pour indiquer que l'ordre entre ces produits n'est pas imposé. Ainsi, {A,E}BCD représente les deux séquences possibles : AEBCD et EABCD.

3. *Croissance des séquences par génération de candidats*

Durant cette phase, des candidats sont créés en faisant croître la longueur des séquences fréquentes grâce aux différents éléments de l'ensemble des produits fréquemment achetés. Ces candidats sont ensuite testés pour vérifier leur existence au sein d'au moins deux suites d'achats.

Les séquences ainsi obtenues sont données dans le Tableau 5 :

Tableau 5 : Séquences fréquentes de longueur 2 (inspiré de (Agrawal & Srikant 1995))

Séquences fréquentes	Support
AB	2
AC	4
AD	3
AE	3
BC	2
BD	2
CD	3
CE	2
DE	2

Ces séquences de longueur deux sont ensuite utilisées pour générer les candidats des séquences fréquentes de longueur trois (présentées dans le Tableau 6), et de même, celles de longueur trois servent pour les candidats des séquences de longueur quatre (présentées dans

le Tableau 7). Cette démarche itérative continue jusqu'à ce qu'il n'y ait plus de candidats qui se retrouvent dans les suites d'achats. Dans cet exemple, les séquences fréquentes les plus longues sont de longueur quatre.

Tableau 6 : Séquences fréquentes de longueur 3 (inspiré de (Agrawal & Srikant 1995))

Séquences fréquentes	Support
ABC	2
ABD	2
ACD	3
ACE	2
BCD	2

Tableau 7 : Séquences fréquentes de longueur 4 (inspiré de (Agrawal & Srikant 1995))

Séquences fréquentes	Support
ABCD	2

4. *Maximisation des séquences*

Cette phase cherche à supprimer les séquences fréquentes qui peuvent se retrouver au sein d'une séquence plus longue. Le but de cet algorithme étant d'identifier des relations entre différents produits, les séquences élémentaires qui ne contiennent qu'un produit sont également supprimées.

Le résultat obtenu à l'issue de cet algorithme est donné dans le Tableau 8 :

Tableau 8 : Séquences fréquentes maximales (inspiré de (Agrawal & Srikant 1995))

Séquences fréquentes	Support
ABCD	2
ABD	2
ACD	3
ACE	2
AD	3
AE	3
DE	2

Ce tableau montre toutes les séquences fréquentes mais il est possible de réduire la taille de celui-ci. En effet, l'analyse de ABCD incluant l'analyse de ABD, il est inutile d'étudier la

séquence ABD. De manière générale, il est possible de ne conserver que les séquences qui ont des alphabets différents tout en gardant les séquences les plus longues. Ainsi, il n'y aurait plus que les séquences fréquentes ABCD, ACE et DE. Nous les appellons les séquences fréquentes significatives.

Dans le cadre de nos travaux, nous souhaitons identifier toutes les adjacences fréquentes entre alarmes. Nous analysons donc avec l'expert toutes les séquences fréquentes significatives qui existent au sein de la séquence d'alarmes.

c. Algorithme développé

Le problème pour lequel l'algorithme *AprioriAll* a été développé présente une différence majeure par rapport à notre problème de la réduction du flux d'alarmes : cet algorithme a été conçu pour chercher des séquences fréquentes significatives en comparant plusieurs séquences alors que nous modélisons l'historique d'alarmes comme une seule séquence. Il est donc nécessaire de développer un fenêtrage pour adapter notre problème à l'algorithme *AprioriAll*.

Comme indiqué dans (Krishnan & Cook 2014), plusieurs méthodes existent pour fenêtrer une séquence et la découper en plusieurs sous-séquences appelées fenêtres d'observation. La Figure 14 montre plusieurs de ces méthodes appliquées à une séquence d'alarmes :

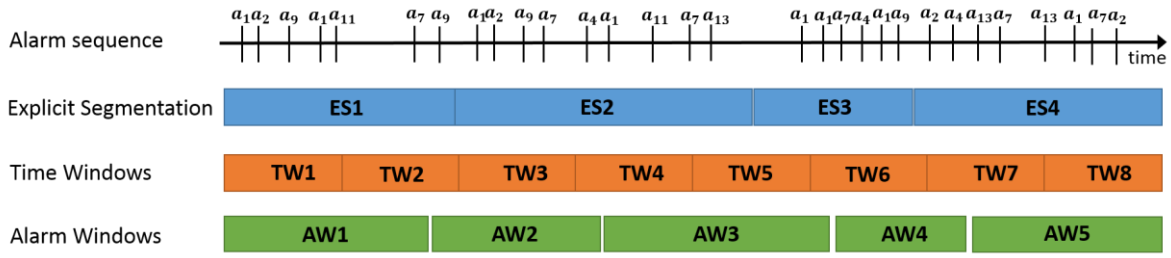


Figure 14 : Approches classiques pour fenêtrer une séquence (inspiré par (Krishnan, et al., 2014))

Les trois fenêtrages indiqués sur la Figure 14 sont les suivants :

- Fenêtrage explicite : Ce fenêtrage est basé sur des connaissances du processus et partitionne la séquence d'alarmes en fonction de phénomènes tel qu'un changement de phase du processus.
- Fenêtrage temporel : Ce fenêtrage est basé sur le temps physique et la séquence d'alarmes est découpée avec un intervalle de temps fixe.
- Fenêtrage d'alarmes : Ce fenêtrage est basé sur le nombre d'occurrences d'alarmes et sépare la séquence d'alarmes en séquences de longueur fixe.

Le fenêtrage que nous choisissons dépend des hypothèses de notre étude. Etant donné que notre méthode n'utilise pas de connaissances du système, le fenêtrage explicite n'est pas possible. De même, nous ne prenons pas en compte le type physique et considérons uniquement l'ordre d'occurrences des alarmes. Il n'est donc pas possible d'utiliser un fenêtrage temporel non plus. Nous optons donc pour une fenêtre glissante de longueur constante en nombre d'occurrences.

Ce type de fenêtre a deux paramètres : la longueur de la fenêtre et le pas avec lequel glisse la fenêtre. La longueur de la fenêtre est le nombre d'occurrences d'alarmes que contient la fenêtre, et le pas de la fenêtre est le décalage (en nombre d'occurrences) qui existe entre la première occurrence d'une fenêtre et la première occurrence de la fenêtre suivante. Cette notion de pas de fenêtrage est représentée sur la Figure 15 :

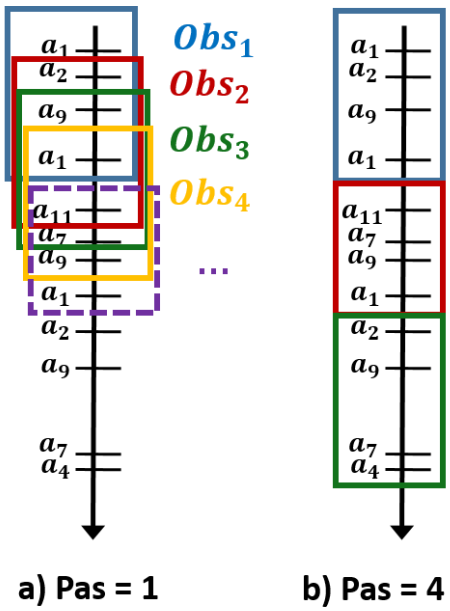


Figure 15 : Exemples de fenêtrage avec une fenêtre de longueur 4

Comme il est possible de le voir sur la Figure 15, une séquence d'alarmes est découpée en fenêtres d'observation de longueur quatre occurrences. Les cas de deux de valeurs de pas sont présentés : un pas de 1 et un pas de 4, égal à la longueur de la fenêtre.

En mettant en parallèle ces deux possibilités, nous faisons ressortir le fait qu'un pas de 1 augmente énormément le nombre de fenêtres d'observation et augmente ainsi le nombre de séquences à analyser. Le second problème avec ce pas est que les fenêtres se recouvrent, ce qui conduit à créer artificiellement des motifs fréquents.

Il est également possible de voir qu'un pas égal à la longueur de la fenêtre assure que les fenêtres d'observations ne se superposent pas et rend ainsi les séquences à analyser indépendantes. Cependant, avec cette solution, les informations situées à la frontière des fenêtres d'observation seront perdues. En effet, une adjacence fréquente pourrait exister entre la dernière occurrence d'une fenêtre et la première de la suivante.

Etant donné que la séquence d'alarmes représente un grand volume de données et contient de nombreuses occurrences (jusqu'à plusieurs centaines de milliers), si un motif fréquent se trouvait coupé au niveau d'une frontière de fenêtres d'observation, ce serait la valeur du support de ce motif qui en serait modifiée. Le risque qu'un motif fréquent soit systématiquement coupé par le fenêtrage, au point que celui-ci ne soit pas détecté par l'algorithme, est très faible.

Le dimensionnement de la fenêtre d'observation est un problème auquel nous n'avons pas apporté de réponse générale, mais auquel nous proposons une discussion lors du traitement industriel d'un exemple dans le Chapitre 5.

Avec ce choix de paramètres, l'algorithme développé peut être représenté tel que ci-dessous :

Algorithme : Identification de motifs fréquents

```
1  Require: Alarm Sequence S
2  windowLength = ??
3  ListWindows = []
4  start = 0
5  Window = []
6  while start + windowLength ≤ |S| do
7      for Alarm in range (start, start + windowLength) do
8          Add Alarm to Window
9      end for
10     Add Window to ListWindows
11     Window = []
12     start = start + windowLength
13 end while
14 Patterns = AprioriAll(ListWindows)
15 return Patterns, List of detected frequent patterns
```

L'algorithme d'identification de motifs fréquents est simple : la première ligne indique l'entrée de cet algorithme tandis que les lignes 2 à 5 initialisent les variables. Ensuite, pour les lignes 6 à 13, la séquence est parcourue pour créer les fenêtres d'observation (lignes 7 à 9) et pour les stocker en entrée de l'algorithme *AprioriAll* (ligne 10). Une fois cette étape franchie, l'algorithme *AprioriAll* est exécuté avec la liste des fenêtres d'observation en entrée (ligne 14). Enfin, la ligne 15 est ici pour indiquer que l'algorithme retourne la liste des motifs fréquents détectés en sortie.

Etant donné que la recherche de motifs fréquents ne se fait plus sur l'entièreté de la séquence d'alarmes mais sur de multiples fenêtres d'observation, la définition du support que nous utilisons a changé :

Définition 2 bis : *Support d'un motif* :

Le support d'un motif est le nombre de fenêtres d'observation Obs_i dans lequel ce dernier a été identifié, et est noté $Supp()$.

Un exemple de résultat de l'algorithme *AprioriAll* adapté à notre problème est donné sur la Figure 16 :



Figure 16 : Exemple simple d'application de l'algorithme AprioriAll

Avec ces trois fenêtres d'observation de longueur 4, les seules séquences qui se retrouvent dans plusieurs fenêtres sont a_1a_2 (dans les fenêtres Obs_1 et Obs_2) et a_7a_4 (dans les fenêtres Obs_2 et Obs_3). Ces deux séquences fréquentes sont présentes dans seulement deux fenêtres d'observation. Leur valeur de support est donc 2 pour toutes les deux.

La démarche complète de la méthode développée ainsi que l'analyse experte réalisée sur les motifs fréquents sont présentées sur un exemple plus complexe.

3) Présentation de l'exemple

L'exemple choisi pour appliquer notre approche est un système de production d'énergie présenté dans (Mikkonen et al. 2016) :

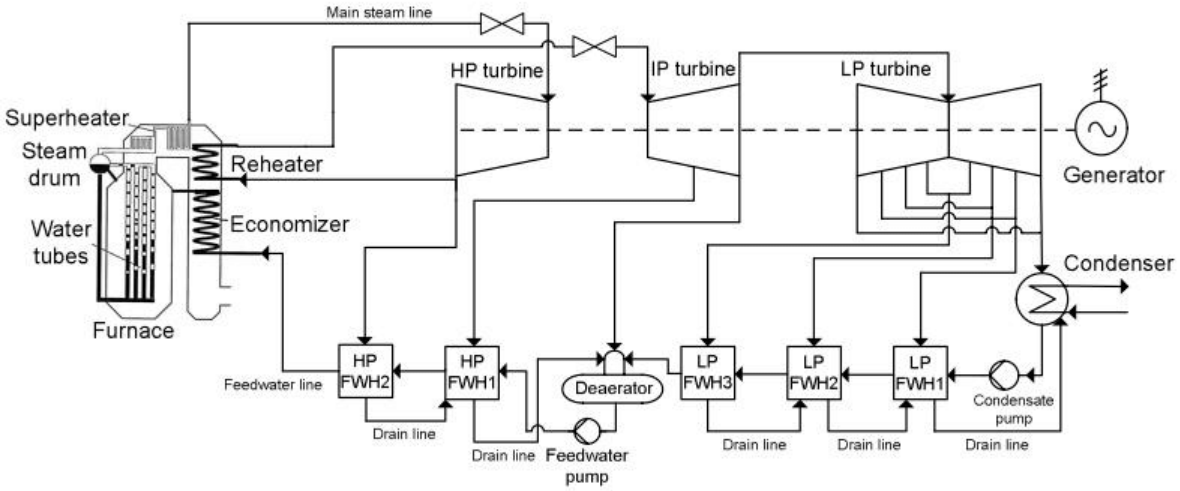


Figure 17 : Schéma représentant un système de production d'énergie (Mikkonen, et al., 2016)

La production d'électricité est basée sur un cycle thermodynamique de Rankine (régénératif et réchauffé). Sur toute centrale à vapeur, l'eau d'alimentation est préchauffée (FWH : *FeedWater Heaters*) à basse pression (BP) et à haute pression (HP) par extraction de la vapeur des turbines. L'efficacité du cycle de vapeur s'en trouve améliorée, mais cela réduit la puissance de sortie de l'installation.

Une chaudière à vapeur typique telle que celle apparaissant sur la Figure 17 est composée d'un foyer, d'un économiseur, de tubes d'eau, d'une cuve, de surchauffeurs et de réchauffeurs. Le carburant est brûlé dans le foyer avec l'excès d'air et une partie de l'eau d'alimentation est évaporée dans les tubes d'eau. La vapeur saturée est séparée de l'eau dans une cuve à vapeur et elle est surchauffée avant la turbine afin d'éviter une condensation dans la turbine.

Le cas d'étude que nous considérons se limite à la chaudière et à son environnement immédiat. Ainsi, sont considérés les alarmes configurées suivantes :

- a_1 : Pression haute dans la cuve de vapeur ;
- a_2 : Pression très haute dans la cuve de vapeur ;
- a_3 : Débit de combustible trop faible ;
- a_4 : Niveau de CO_2 trop élevé ;
- a_5 : Débit trop important en sortie de la cuve de vapeur ;

- a_6 : Vanne d'alimentation en combustible mal ouverte ;
- a_7 : Vanne d'alimentation en combustible mal fermée ;
- a_8 : Défaut d'alimentation électrique de la chaudière ;
- a_9 : Température paroi extérieure trop élevée ;
- a_{10} : Température surchauffeur trop faible ;

On considère également que l'historique du système d'alarmes a permis d'obtenir la séquence d'alarmes suivante :

$$S = a_1a_3a_1a_8a_9a_2a_5a_6a_4a_3a_1a_2a_5a_{10}a_7a_{10}a_4a_1a_2a_5a_7a_4a_9$$

$$a_6a_9a_3a_4a_7a_8a_1a_9a_2a_5a_8a_8a_7a_3a_6a_{10}a_3a_{10}a_6a_9a_1a_3a_2a_5$$

4) Application de la méthode sur l'exemple

a. Obtention des motifs fréquents

La séquence d'alarmes ci-dessus sur laquelle nous appliquons notre démarche de recherche d'adjacences fréquentes pour réduire le flux d'alarmes est relativement courte (47 occurrences d'alarmes). De ce fait, le risque de perdre des informations à cause du découpage en fenêtres d'observation avec un pas égal à la longueur de la fenêtre n'est plus négligeable.

Pour notre cas d'application, la longueur des fenêtres d'observation est arbitrairement prise égale à 7 occurrences. Le pas de fenêtrage est donc de 7 lui aussi.

Avec ces paramètres, la séquence d'alarmes S est découpé en sept fenêtres Obs_i :

- $Obs_1 = a_1a_3a_1a_8a_9a_2a_5$;
- $Obs_2 = a_6a_4a_3a_1a_2a_5a_{10}$;
- $Obs_3 = a_7a_{10}a_4a_1a_2a_5a_7$;
- $Obs_4 = a_4a_9a_6a_9a_3a_4a_7$;
- $Obs_5 = a_8a_1a_9a_2a_5a_8a_8$;
- $Obs_6 = a_7a_3a_6a_{10}a_3a_{10}a_6$;
- $Obs_7 = a_9a_1a_3a_2a_5$;

Ces fenêtres sont ensuite analysées par l'algorithme *AprioriAll* et les quatre motifs fréquents ainsi identifiés sont donnés dans le Tableau 9 :

Tableau 9 : Motifs fréquents identifiés par AprioriAll

Nom du motif	Motif frequent identifié	Support
P_1	a_2a_5	4
P_2	a_3a_1	2
P_3	$a_1a_2a_5$	2
P_4	$a_9a_2a_5$	2

A ce niveau, il est important de remarquer qu'une alarme peut appartenir à plusieurs motifs fréquents. Sachant que chacun de ces motifs traduit une relation d'adjacence fréquente, il peut donc être préférable de considérer simultanément plusieurs relations au moment de chercher $\tilde{S}$ telle que $(S,R) \equiv (\tilde{S},R)$ avec $R = \{P_1, P_2, P_3, P_4\}$ l'ensemble des motifs fréquents identifiés.

b. La détermination des séquences équivalentes

Les séquences $\tilde{S}$ qui respectent $(S,R) \equiv (\tilde{S},R)$ et qui peuvent être obtenues depuis S par suppression d'occurrences d'alarmes sont données à l'expert pour qu'il puisse choisir la séquence qui préserve la sûreté du système.

Parmi ces séquences, il est possible de retrouver $\widetilde{S_1}$, $\widetilde{S_2}$, $\widetilde{S_3}$ et $\widetilde{S_4}$:

- $\widetilde{S_1} = a_1a_3a_1a_8a_9a_2a_6a_4a_3a_1a_2a_{10}a_7a_{10}a_4a_1a_2a_7a_4a_9a_6a_9a_3a_4$
  $a_7a_8a_1a_9a_2a_8a_8a_7a_3a_6a_{10}a_3a_{10}a_6a_9a_1a_3a_2$
- $\widetilde{S_2} = a_1a_3a_8a_9a_2a_5a_6a_4a_3a_2a_5a_{10}a_7a_{10}a_4a_1a_2a_5a_7a_4a_9a_6a_9a_3$
  $a_4a_7a_8a_1a_9a_2a_5a_8a_8a_7a_3a_6a_{10}a_3a_{10}a_6a_9a_1a_3a_2a_5$
- $\widetilde{S_3} = a_1a_3a_8a_9a_2a_6a_4a_3a_2a_{10}a_7a_{10}a_4a_1a_2a_7a_4a_9a_6a_9a_3a_4a_7a_8$
  $a_1a_9a_2a_8a_8a_7a_3a_6a_{10}a_3a_{10}a_6a_9a_1a_3a_2$
- $\widetilde{S_4} = a_3a_2$

La séquence $\widetilde{S_1}$ est obtenue en supprimant les occurrences de a_5 qui suivent des occurrences de a_2 et respecte $(S,R) \equiv (\widetilde{S_1},R)$. Il possible de vérifier que cette séquence est bien équivalente en testant pour les différents motifs l'existence de séquences respectant la définition de l'équivalence présentée page 45. La vérification du respect de la relation d'adjacence fréquente indiquée par P_1 est donnée en exemple :

Soit $\Sigma_r = \{a_2, a_5\}$ l'alphabet de projection : $Proj_{\Sigma_r}(S) = a_2a_5\boldsymbol{a_2a_5}a_2a_5a_2a_5a_2a_5$ et $Proj_{\Sigma_r}(\widetilde{S_1}) = a_2\boldsymbol{a_2}a_2a_2a_2$, il existe donc bien $\tilde{\sigma} = a_2$ tel que $\sigma = a_2a_5$ puisse s'écrire de la forme $\sigma = \varepsilon\tilde{\sigma}a_5$ avec $(\varepsilon, a_5) \in (\Sigma_r{}^*)^2$.

De même, la séquence $\widetilde{S_2}$ est obtenue en supprimant les occurrences de a_1 qui suivent des occurrences de a_3 et respecte $(S, R) \equiv (\widetilde{S_2}, R)$. $\widetilde{S_3}$ est construite en supprimant les occurrences de S qui ont été supprimées pour obtenir $\widetilde{S_1}$ et $\widetilde{S_2}$, et $\widetilde{S_4}$ est une séquence qui respecte nos critères d'équivalence mais que l'expert va facilement pouvoir écarter. L'expert va rejeter cette solution car la conservation de la sûreté du système ne peut pas être assurée quand nous supprimons des alarmes n'appartenant à aucune relation.

c. Analyse experte

Pour illustrer l'utilisation de notre méthode, nous détaillons un exemple d'analyse qui peut être effectuée par l'expert à partir des motifs fréquents identifiés et des séquences équivalentes proposées. Nous nous limitons à la prise en compte du seul motif fréquent P_1 car, lorsque l'expert effectue son analyse, il commence par les motifs qui ont les plus gros supports, ceux-ci permettant potentiellement une réduction plus importante du flux d'alarmes. Le motif a_2a_5 a un support de 4 car il a été identifié dans les fenêtres d'observation Obs_1, Obs_2, Obs_3 et Obs_7.

a_2 correspond à l'alarme « pression très haute dans la cuve » et a_5 correspond à l'alarme « débit trop élevé en sortie de cuve ». En considérant la configuration de ces alarmes sur la chaudière, le débit est directement lié à la pression dans la cuve. L'opérateur reçoit donc des informations redondantes et doit effectuer la même action « ouvrir la vanne de surpression de la cuve » pour répondre à ces deux alarmes. Il est donc possible de ne garder qu'une de ces alarmes. En choisissant de conserver la cause plutôt que la conséquence, les occurrences de a_5 qui suivent les occurrences de a_2 peuvent être supprimés.

Cette analyse permet ainsi de valider la séquence $\widetilde{S_1}$ comme séquence équivalente permettant de réduire le flux d'alarmes tout en conservant la sûreté du système. Avec ce choix de séquences, la réduction obtenue sur la longueur de la séquence et donc le flux d'alarmes est de 10,6%.

Pour obtenir la séquence $\tilde{S}$ équivalente à la séquence d'alarmes S qui prend en compte tous les motifs fréquents, l'expert doit analyser tous ces motifs. Or la recherche d'adjacences fréquentes donne un grand nombre de motifs fréquents. Tous les analyser ne serait pas rentable pour l'expert étant donné que la majorité des motifs fréquents identifiés par cette

analyse possède un petit support et, par conséquent, que la réduction éventuelle de la longueur de S serait donc minime.

Ainsi, dans notre cas d'étude, l'expert juge que le support de 2 des motifs fréquents restants n'est pas suffisant pour justifier une analyse et que ces motifs ne sont pas représentatifs du comportement des alarmes. A la fin de l'étude, la séquence $\tilde{S}$ équivalente à la séquence d'alarmes S est donc égale à $\widetilde{S_1}$.

5) Conclusions

La méthode détaillée dans ce chapitre a été présentée dans (Laumonier et al. 2017) et permet de réduire le flux d'alarmes arrivant aux opérateurs grâce à l'identification de toutes les adjacences fréquentes qui se traduisent par des motifs fréquents. Cependant, cette force est également sa faiblesse car identifier tous les motifs fréquents demande un temps de calcul très long.

De plus, cette méthode se limite à l'identification de relations entre occurrences d'alarmes adjacentes. C'est pour dépasser cette limitation que nous avons développé une seconde méthode utilisée pour identifier des précédences systématiques entre des occurrences qui ne sont pas forcément adjacentes.

Chapitre 4 : Identification de précédences systématiques entre alarmes

Dans ce chapitre, nous présentons une deuxième méthode servant à identifier des précédences systématiques entre alarmes. Ce sont des précédences telles que les occurrences d'une alarme sont toujours précédées de celles d'une autre sans qu'il y ait forcément adjacence. Nous catégorisons ces relations en deux types appelés domination et mutuelle dépendance.

Etant donné que la méthode utilisant les adjacences fréquentes fournit un très grand nombre de séquences équivalentes potentielles, l'expert passe beaucoup de temps à analyser ces résultats. Pour éviter cette limitation, nous souhaitons faciliter et guider son analyse en représentant ces précédences dans un réseau de Petri.

Pour illustrer cette méthode, nous l'appliquons au cas de la chaudière à vapeur présenté dans le chapitre 3.

1) Réduction de la séquence par analyse de précédences systématiques

L'objectif de la méthode développée dans ce chapitre est toujours de proposer à l'expert une séquence équivalente qui permettrait de réduire le flux d'alarmes tout en assurant la sûreté du système industriel. La différence est que, à la place d'analyser les adjacences fréquentes, l'expert va choisir la séquence qui respecte ces critères en analysant les précédences systématiques qui existent dans la séquence d'alarmes initiale.

a. Analyse sur des séquences projetés

L'hypothèse qui a guidé notre travail est que s'il existe une précédence systématique entre deux alarmes, alors cette relation n'est pas conditionnée par d'autres alarmes. Avec cette considération, la recherche des précédences systématiques peut être effectuée en considérant les alarmes deux à deux.

Ainsi, l'idée est de ne pas rechercher les précédences systématiques sur S dans son entièreté, mais de rechercher ce type de précédences sur des projections de S sur des alphabets formés de deux alarmes différentes.

b. Obtenir la séquence équivalente

Une alarme peut intervenir dans plusieurs précédences systématiques. Lorsque ces relations ont des alarmes en commun, nous disons que ces relations font partie de la même famille de précédences systématiques. Ces différentes familles sont donc complètement isolées et possèdent des alphabets d'alarmes disjoints. Ainsi, l'analyse effectuée sur une famille est indépendante de l'analyse effectuée sur d'autres familles. De plus, la réduction de la longueur

totale de S peut être vue comme de multiples réductions de la longueur de projections de S sur des alphabets disjoints.

Ces familles représentant les relations de précédences systématiques, si F est l'ensemble des familles f_i, l'expert doit choisir une séquence équivalente $\tilde{S}$ parmi celles qui respectent $(S, F) \equiv (\tilde{S}, F)$ et $|\tilde{S}| < |S|$.

Il est donc possible d'effectuer l'analyse en considérant indépendamment les différentes familles f_i et en concluant sur la potentielle suppression d'occurrences des alarmes appartenant à l'alphabet des alarmes de f_i, noté Σ_i. Cette démarche doit permettre d'obtenir des séquences équivalentes $\widetilde{S_i}$ qui respectent la relation suivante :

$$\widetilde{S_i} \in \Sigma_i^+ \mid \begin{cases} (Proj_{\Sigma_i}(S), \{f_i\}) \equiv (\widetilde{S_i}, \{f_i\}) \\ |\widetilde{S_i}| \leq |Proj_{\Sigma_i}(S)| \end{cases}$$

Pour finir, les séquences équivalentes à la séquence d'alarmes S, notée $\tilde{S}$, qui vont être confiées à un expert sont obtenues en utilisant la connaissance des différents $\widetilde{S_i}$:

$$\forall\, \widetilde{S_i}, \begin{cases} Proj_{\Sigma_i}(\tilde{S}) = \widetilde{S_i} \\ \tilde{S} \text{ obtenue par suppression d'occurrences de } S \end{cases}$$

2) Obtenir les précédences systématiques

Les relations de précédences systématiques entre alarmes peuvent se décomposer en deux cas : le cas où l'occurrence d'une alarme est une condition nécessaire pour l'occurrence d'une autre alarme, et le cas où l'occurrence d'une alarme est une condition nécessaire et suffisante pour l'occurrence d'une autre alarme. Le cas où l'occurrence n'est pas nécessaire ne concerne pas la recherche de précédences systématiques.

Ces deux cas possibles portent respectivement le nom de domination et de mutuelle dépendance. Ces notions sont apparues pour la première fois dans (Saives et al. 2015) dans le cadre de la recherche du comportement non observable d'une chaîne de production à partir de l'observation de ses variables entrées/sorties. Nous avons formalisé ces notions dans (Laumonier et al. 2019) pour les adapter à notre problème.

a. Relation de domination

Définition 1 : *Relation de domination* :

Soit une séquence d'alarme S définissant un alphabet Σ_S, et a_i et a_j $(i \neq j)$ deux alarmes différentes de Σ_S. a_i domine a_j, noté $a_i \rightarrow a_j$, si et seulement si la projection de S sur $\{a_i, a_j\}$

est une alternance entre une séquence non nulle d'occurrences de a_i et une unique occurrence de a_j :

$$Proj_{\{a_i,a_j\}}(S) = \sigma_{i,1}a_j\sigma_{i,2}a_j\sigma_{i,3}a_j\sigma_{i,4}\ldots \qquad \text{avec } \forall\, k \in \mathbb{N}^*, \begin{cases}\sigma_{i,k} \in \{a_i\}^* \\ |\sigma_{i,k}| \geq 1\end{cases}$$

Un exemple de domination est présenté dans la Figure 18 :

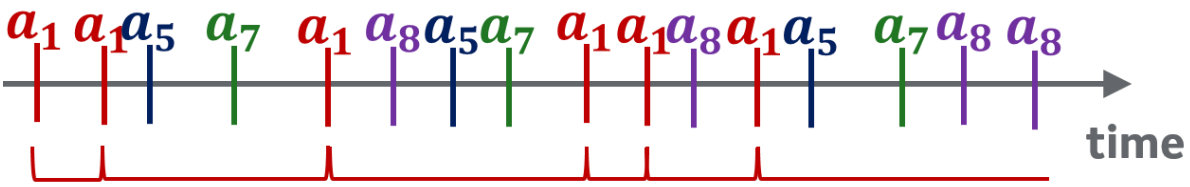


Figure 18 : Exemple de domination dans une séquence d'alarmes

En considérant les alarmes a_1 et a_5, nous retrouvons une séquence projetée $Proj_{\{a_1,a_5\}}(S) = a_1a_1a_5a_1a_5a_1a_1a_1a_5$ qui est une alternance entre une séquence non vide d'occurrences de a_1 et une unique occurrence de a_5. L'alarme a_1 domine donc l'alarme a_5 : $a_1 \rightarrow a_5$. Cela se traduit sur la Figure 18 par le fait que l'alarme a_5 n'a au maximum qu'une occurrence entre deux occurrences de a_1. De même, il est possible de trouver que a_1 domine a_7 et que a_5 domine a_7.

Le fait qu'une alarme a_i domine une alarme a_j indique qu'avant chaque occurrence de a_j au moins une occurrence de a_i a été observé. Le fait que l'alarme a_j apparaisse implique donc que l'alarme a_i a eu lieu auparavant. Cette connaissance offre la possibilité de supprimer dans la séquence d'alarmes les occurrences de a_i qui précèdent immédiatement les occurrences de a_j dans $Proj_{\{a_i,a_j\}}(S)$. C'est cette considération qui est utilisée pour déterminer les séquences équivalentes qui seront confiées à l'expert.

b. Relation de mutuelle dépendance

Définition 2 : *Relation de mutuelle dépendance* :

Soit une séquence d'alarme S définissant un alphabet Σ_S, et a_i et a_j deux alarmes différentes de Σ_S. a_i et a_j sont mutuellement dépendantes, noté $a_i \leftrightarrows a_j$, si et seulement si la projection de S sur $\{a_i, a_j\}$ est une alternance entre une unique occurrence de a_i et une unique occurrence de a_j :

$$\begin{cases} Proj_{\{a_i,a_j\}}(S) = a_ia_ja_ia_ja_i\ldots \\ \qquad\qquad ou \\ Proj_{\{a_i,a_j\}}(S) = a_ja_ia_ja_ia_j\ldots \end{cases}$$

Un exemple de mutuelle dépendance est présenté dans la Figure 19 :

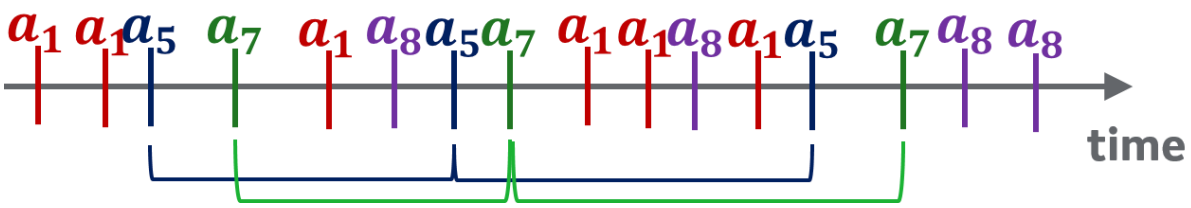


Figure 19 : Exemple de mutuelle dépendance dans une séquence d'alarmes

En considérant les alarmes a_5 et a_7, nous obtenons une séquence projetée $Proj_{\{a_5,a_7\}}(S) = a_5a_7a_5a_7a_5a_7$ qui est une alternance entre une unique occurrence de a_5 et une unique occurrence de a_7. Les alarmes a_5 et a_7 sont donc mutuellement dépendantes : $a_5 \leftrightarrows a_7$. Cette relation se traduit sur la Figure 19 par le fait que l'alarme a_7 n'a qu'une occurrence entre deux occurrences de a_5, et vice-versa.

La relation de domination entre a_5 et a_7 qui a été identifiée précédemment est incluse dans cette relation de mutuelle dépendance et n'est donc pas utilisée dans le reste de l'étude.

Le fait que deux alarmes a_i et a_j soient mutuellement dépendantes indique qu'avant chaque occurrence de a_j (ou de a_i) une occurrence de a_i (ou de a_j) a été observée. Cette observation implique que la connaissance de l'occurrence d'une des deux alarmes suffit pour savoir que la deuxième a eu lieu. Cette dernière offre la possibilité de supprimer dans la séquence d'alarmes les occurrences d'une de ces deux alarmes. Un choix possible est de garder la première des deux alarmes qui apparaît dans la séquence est de supprimer les occurrences de la seconde. Cependant, d'autres possibilités existent et ce sont ces différentes possibilités qui sont considérées pour déterminer les différentes séquences équivalentes qui seront confiées à l'expert.

Les relations de domination et de mutuelle dépendance qui existent au sein de la séquence d'alarmes sont obtenues par un algorithme qui parcourt la séquence pour chaque couple d'alarmes en vue de vérifier l'existence de précédences systématiques entre ces deux alarmes. Cet algorithme est relativement simple et présente une complexité théorique polynomiale de $\mathcal{O}(|\Sigma_S|^2|S|)$ où S est la séquence d'alarmes et Σ_S est l'alphabet d'alarmes.

Pour regrouper des alarmes qui appartiendraient à plusieurs relations, et pour ainsi aider l'expert dans son analyse, nous avons choisi de représenter ces relations de manière graphique. Dans le graphe ainsi obtenu, quand des relations existent entre des alarmes, ces dernières se retrouvent groupées au sein d'un même fragment connexe du graphe global. Ce choix présente

l'avantage d'isoler des alarmes ou groupements d'alarmes indépendants les uns des autres, et d'ainsi encore plus faciliter l'analyse experte.

3) Représenter les précédences systématiques

Pour traduire les dominations et les mutuelles dépendances de manière graphique tout en conservant le formalisme de notre méthode, nous avons utilisé une classe de réseaux de Petri (RdP) comme représentation graphique. Pour construire cette représentation, nous avons réalisé une bijection entre l'alphabet des alarmes et l'ensemble des transitions. Dans un souci de clarté, les transitions portent le nom des alarmes.

L'approche que nous proposons dans cette thèse est soumise à des hypothèses qui conditionnent la classe de RdP que nous pouvons utiliser. Ainsi, l'absence de connaissance sur le procédé et la non prise en compte du temps physique incitent à choisir la classe des RdP autonomes.

Définition 3 : *Réseau de Petri autonome* (Murata 1989) :

Un réseau de Petri (RdP) autonome est un graphe bipartite possédant deux types de sommets : les places et les transitions. Les places sont représentées par des cercles et les transitions par des barres. Des arcs orientés relient les places aux arcs et vice-versa.

Formellement, un RdP est un 4-uplet : $PN = (P, T, I, O)$ où :

- $P = \{p_1, p_2, \dots, p_{|P|}\}$ est un ensemble fini de places ;
- $T = \{t_1, t_2, \dots, t_{|T|}\}$ est un ensemble fini de transitions ;
- $\begin{cases} P \cap T = \emptyset \\ P \cup T \neq \emptyset \end{cases}$
- $I : P \times T \rightarrow \{0,1\}$ est une fonction représentant l'existence d'arcs allant de places vers des transitions ;
- $O : T \times P \rightarrow \{0,1\}$ est une fonction représentant l'existence d'arcs allant de transitions vers des places ;

Une fonction marquage $M : P \rightarrow Z^+$ associe à chaque place de P un nombre entier de jetons. Le marquage d'un RdP est usuellement exprimé par un vecteur de taille $|P|$. M_0 est le marquage initial d'un RdP marqué $PN = (P, T, I, O, M_0)$.

Dans un RdP, une transition t_j est tirable pour un marquage M_k si $\forall\, p_i \in P$, $M_k(p_i) \geq I(p_i, t_j)$. En tirant cette transition, un nouveau marquage M_{k+1} est atteint. Celui-ci peut être calculé par $M_{k+1} = M_k + (O - I)u_k$, où $u_k(i) = 0, i \neq j$, et $u_k(j) = 1$.

L'ensemble de toutes les séquences de tir de transitions qu'il est possible d'obtenir en faisant évoluer le RdP à partir d'un marquage donné s'appelle le langage généré à partir de ce marquage.

Avec notre approche, pour regrouper les relations existantes autour des alarmes, nous proposons un RdP où l'ensemble des transitions T est l'alphabet de la séquence d'alarmes Σ_s. Cette bijection entre T et Σ_s signifie que les alarmes sont représentées comme des transitions et sont uniques dans le RdP.

Le RdP qui traduit les relations entre alarmes est donc composé de fragments connexes dont les alphabets d'alarmes sont disjoints. Ce sont ces fragments qui sont la traduction des familles de précédences systématiques et qui permettent de déterminer des séquences équivalentes.

a. Traduction d'une domination

Une relation de domination représente le fait que l'occurrence d'une alarme a_i est nécessaire pour que la deuxième alarme a_j puisse se produire. Cela signifie donc que la transition a_i doit d'abord être tirée avant que a_j ne soit tirée. Ceci se traduit que par le fait que la transition a_i va générer un jeton qui va être stocké dans une place amont à la transition a_j.

La traduction en RdP de la relation de domination précédente $a_1 \rightarrow a_5$ est donnée sur la Figure 20 :

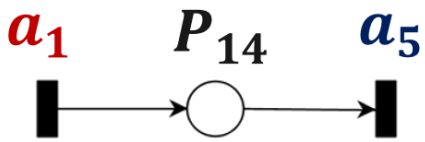


Figure 20 : Représentation d'une relation de domination

Pour que a_5 puisse être tirée dans ce RdP, un jeton doit être présent dans la place P_{14}. Ceci n'est possible que si a_1 a été tirée auparavant.

En faisant vivre ce RdP, il devient évident que ce réseau permet des séquences de tirs qui ne correspondent pas aux relations de dominations identifiées. C'est le cas de la séquence de tir $a_1 a_1 a_5 a_5$ par exemple (la deuxième occurrence de a_5 n'est pas précédée d'au moins une occurrence de a_1). Il faut donc être conscient que cette traduction présente un langage excédentaire et que la séquence d'alarmes n'est qu'une séquence de ce langage généré.

b. Traduction d'une mutuelle dépendance

Une relation de mutuelle dépendance représente le fait que l'occurrence d'une alarme a_i est une condition nécessaire et suffisante pour que la deuxième alarme a_j puisse se produire. Cela signifie donc que la transition a_i doit d'abord être tirée avant que a_j ne soit tirée, et que le tir suivant de a_i ne peux avoir lieu qu'après le tir de a_j. Ceci se traduit que par le fait que la transition a_i va générer un jeton qui va être stocké dans une place amont à la transition a_j et que la transition a_j va elle-même générer un jeton qui sera stocké dans une place amont à a_i. Si la séquence d'alarme projetée sur $\{a_i, a_j\}$ commence par une occurrence de a_i, un jeton est initialement placé dans cette place.

La traduction en RdP de la relation de domination précédente $a_5 \leftrightarrows a_7$ est donnée sur la Figure 21 :

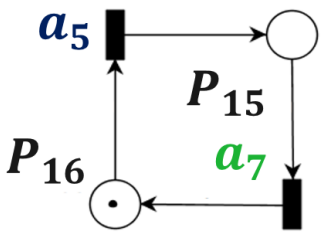


Figure 21 : Représentation d'une relation de mutuelle dépendance

Pour que le tir de a_5 soit possible, la place P_{16} doit posséder un jeton. Quand a_5 est tirée, un jeton est enlevé de P_{16} et un jeton est ajouté dans P_{15}. La place P_{15} possédant désormais un jeton, la transition a_7 peut-être tirée. Ce tir enlève un jeton à P_{15} et rajoute un jeton à P_{16}, autorisant à nouveau le tir de a_5.

Etant donné que $Proj_{\{a_5,a_7\}}(S) = a_5a_7a_5a_7a_5a_7$, la première occurrence est une occurrence de a_5. La première transition qui doit être tirée est donc a_5. Le marquage initial de ce RdP est un jeton dans P_{16} et zéro jeton dans P_{15}.

A la différence de la relation de domination, ce réseau ne présente pas de langage excédentaire et son évolution permet donc de visualiser directement la manière dont interagissent les alarmes mutuellement dépendantes.

c. Assemblage des fragments

Comme il a été dit précédemment, le choix a été fait de représenter une alarme par une unique transition. Ainsi, si une alarme est impliquée dans plusieurs relations de domination

et/ou de mutuelle dépendance, les fragments associés à ces relations vont être assemblés au niveau de leur transition commune pour ne former qu'un seul fragment connexe de RdP.

Un exemple de cet assemblage est donné sur la Figure 22 pour la domination $a_1 \rightarrow a_5$ et la mutuelle dépendance $a_5 \leftrightarrows a_7$:

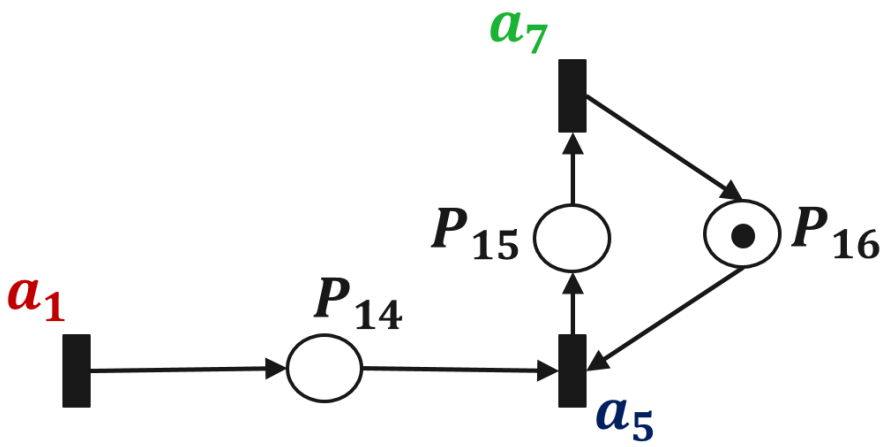


Figure 22 : Assemblage des relations au niveau des transitions communes

La domination $a_1 \rightarrow a_5$ et la mutuelle dépendance $a_5 \leftrightarrows a_7$ possèdent l'alarme a_5 en commun. Le RdP obtenu est donc la fusion des réseaux de la Figure 20 et de la Figure 21 au niveau de la transition a_5.

Le réseau représentant une relation de domination ayant un langage généré excédentaire, tout réseau qui combine au moins une relation de domination à d'autres relations va lui aussi posséder un langage excédentaire.

En appliquant cette démarche sur l'entièreté des dominations et mutuelles dépendances trouvées durant l'analyse de la séquence exemple $S = a_1a_1a_5a_7a_1a_8a_5a_7a_1a_1a_8a_1a_5a_7a_8a_8$, qui sont $a_1 \rightarrow a_5$, $a_1 \rightarrow a_7$ et $a_5 \leftrightarrows a_7$, nous obtenons le RdP de la Figure 23 :

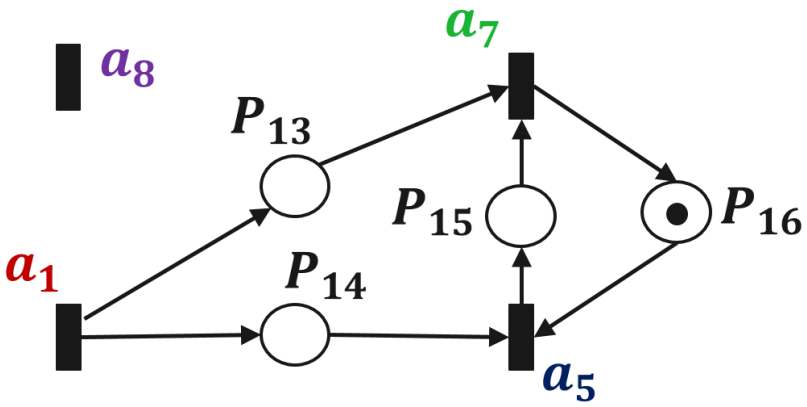


Figure 23 : Représentation complète des relations entre alarmes

Nous remarquons que l'alarme a_8 qui n'appartient à aucune relation de précédence systématique apparaît également dans ce réseau. Ce choix est fait car nous souhaitons obtenir un RdP dont le langage généré contient la séquence d'alarmes S. De plus, le fait que notre méthode ait identifié a_8 comme une alarme indépendante des autres est une information utile pour l'obtention des séquences équivalentes. En effet, si une alarme est indépendante des autres alarmes, elle transporte une information qui lui est propre et ne doit donc pas avoir d'occurrences supprimées.

d. Réduction des places implicites

La méthode de traduction et d'assemblage des dominations et des mutuelles dépendances est formelle et permet de retrouver toutes ces relations dans un RdP. Cependant, cette méthode traite de manière indépendante les différentes relations lors de leur traduction avant leur assemblage. Ainsi, l'assemblage des différents fragments génère des places implicites au sein du RdP.

Définition 4 : *Places implicites* (Silva 1993) :

Une place implicite est une place dont la suppression ne change rien à l'ensemble des marquages accessibles du RdP. C'est-à-dire que la suppression d'une place implicite ne change pas le langage généré du RdP.

Pour faciliter au maximum la lecture du RdP, et ainsi faciliter l'analyse de l'expert, nous cherchons donc à supprimer les places implicites présentes dans le réseau correspondant à l'assemble brut des relations.

Le problème de la détermination des places implicites est un problème d'optimisation linéaire. L'algorithme que nous avons utilisé est celui qui est proposé dans (Silva 1993) et qui possède une complexité polynomiale.

En appliquant cette démarche sur le réseau complet brut de la Figure 23, et en supprimant la place P_{13} qui est la seule place implicite du réseau, nous obtenons le RdP complet nettoyé des places implicites qui est représenté sur la Figure 24 :

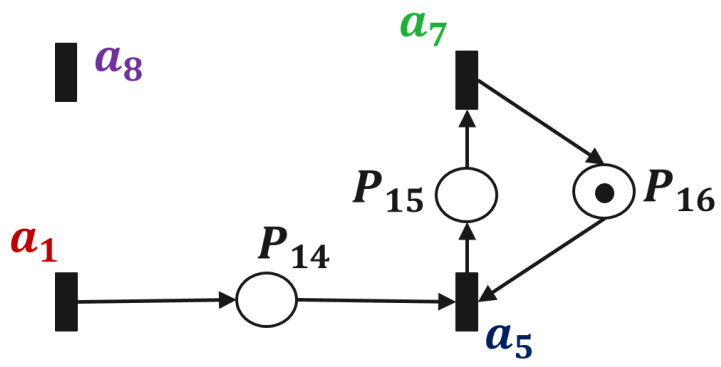


Figure 24 : Représentation complète des relations entre alarmes avec suppression des places implicites

4) Application au cas de la chaudière

Nous appliquons notre démarche sur le cas d'étude de la chaudière à vapeur présenté dans le chapitre 3 :

$$S = a_1a_3a_1a_8a_9a_2a_5a_6a_4a_3a_1a_2a_5a_{10}a_7a_{10}a_4a_1a_2a_5a_7a_4a_9$$

$$a_6a_9a_3a_4a_7a_8a_1a_9a_2a_5a_8a_8a_7a_3a_6a_{10}a_3a_{10}a_6a_9a_1a_3a_2a_5$$

a. Le réseau de Petri traduisant les précédences systématiques entre alarmes

La recherche de dominations et de mutuelles dépendances a permis de découvrir les relations suivantes :

- $a_1 \rightarrow a_2$;
- $a_1 \rightarrow a_5$;
- $a_2 \rightarrow a_5$;
- $a_3 \rightarrow a_6$;
- $a_2 \leftrightarrows a_5$ (la domination $a_2 \rightarrow a_5$ est incluse dans cette mutuelle dépendance) ;

Les cinq alarmes a_4, a_7, a_8, a_9 et a_{10} n'interviennent dans aucune relation de domination ou de mutuelle dépendance et sont donc catégorisées comme indépendantes.

Le réseau de Petri complet et nettoyé des places implicites qui correspond à ces relations est donné sur la Figure 25 :

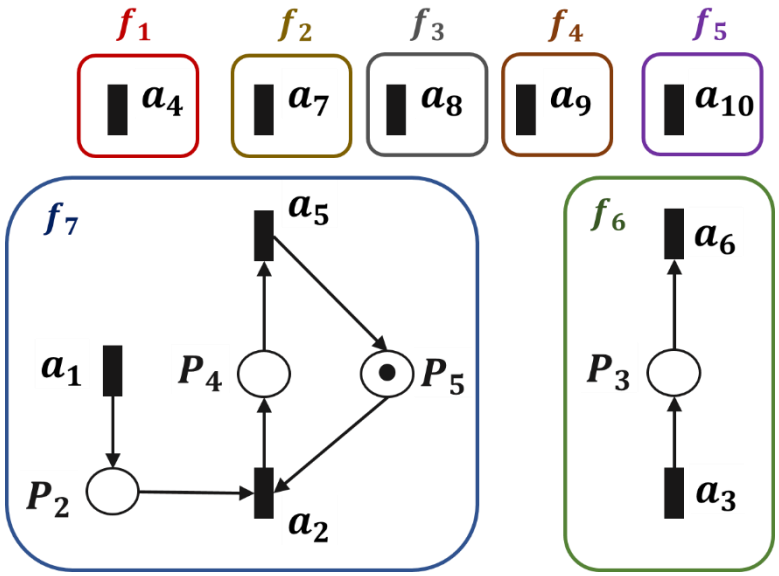


Figure 25 : RdP traduisant les relations identifiées sur l'exemple de la chaudière

Pour obtenir ce réseau, la seule place implicite qui a été supprimée est la place qui reliait les transitions a_1 et a_5.

b. Les séquences équivalentes à chaque fragment

Pour chaque fragment f_i apparaissant sur la Figure 25, nous recherchons toutes les séquences $\widetilde{S_\iota}$ équivalentes à la projection de S sur l'ensemble des transitions du fragment Σ_i :

$$\widetilde{S_\iota} \in \Sigma_i^+ \mid \begin{cases} (Proj_{\Sigma_i}(S), \{f_i\}) \equiv (\widetilde{S_\iota}, \{f_i\}) \\ |\widetilde{S_\iota}| \leq |Proj_{\Sigma_i}(S)| \end{cases}$$

1. *Le cas des alarmes indépendantes*

Pour les cinq fragments correspondant aux cinq alarmes jugées indépendantes (fragments f_1, f_2, f_3, f_4 et f_5), notre méthode recommande de ne supprimer aucune occurrence :

- $\widetilde{S_1} = Proj_{\{a_4\}}(S)$;
- $\widetilde{S_2} = Proj_{\{a_7\}}(S)$;
- $\widetilde{S_3} = Proj_{\{a_8\}}(S)$;
- $\widetilde{S_4} = Proj_{\{a_9\}}(S)$;
- $\widetilde{S_5} = Proj_{\{a_{10}\}}(S)$;

Ces cinq séquences respectent de manière triviale les critères de la définition de séquences équivalentes présentée page 45.

2. *Le cas du fragment composé de a_3 et a_6*

Le fragment f_6 représente une relation de domination entre les alarmes a_3 et a_6. Dans le cas d'une domination, étant donné que l'occurrence d'une alarme dominée implique que l'occurrence de l'alarme dominante a eu lieu, notre méthode recommande de supprimer l'occurrence de l'alarme dominante quand elle précède une occurrence de l'alarme dominée. Dans notre cas, cela revient à supprimer les occurrences de a_3 qui précèdent une occurrence de a_6.

Etant donné que $Proj_{\{a_3,a_6\}}(S) = a_3a_6a_3a_6a_3a_3a_6a_3a_6a_3$, la séquence proposée $\widetilde{S_6}$ est la suivante :

$$\widetilde{S_6} = \cancel{a_3}a_6\cancel{a_3}a_6a_3\cancel{a_3}a_6\cancel{a_3}a_6a_3 = a_6a_6a_3a_6a_6a_3$$

$Proj_{\{a_3,a_6\}}\left(Proj_{\{a_3,a_6\}}(S)\right) = a_3a_6a_3a_6a_3\boldsymbol{a_3a_6}a_3a_6a_3$ et $Proj_{\{a_3,a_6\}}(\widetilde{S_6}) = a_6a_6a_3\boldsymbol{a_6}a_6a_3$, il existe donc bien un $\tilde{\sigma} = a_6$ tel que $\sigma = a_3a_6$ puisse s'écrire de la forme $\sigma = a_3\tilde{\sigma}\varepsilon$ avec $(a_3, \varepsilon) \in (\{a_3, a_6\}^*)^2$. Nous avons donc bien $(Proj_{\Sigma_6}(S), \{f_6\}) \equiv (\widetilde{S_6}, \{f_6\})$. De plus, $|\widetilde{S_6}| < |Proj_{\Sigma_6}(S)|$.

$\widetilde{S_6}$ respecte donc les conditions de la définition des séquences équivalentes.

3. *Le cas du fragment composé de a_1, a_2 et a_5*

Le fragment f_7 représente des relations de domination et une relation de mutuelle dépendance : $a_1 \rightarrow a_2$, $a_1 \rightarrow a_5$ et $a_2 \leftrightarrows a_5$. Dans le cas d'une domination, notre méthode recommande de supprimer l'occurrence de l'alarme dominante quand elle précède une occurrence de l'alarme dominée. De plus, étant donné que dans le cas d'une mutuelle dépendance l'occurrence d'une des deux alarmes implique que l'occurrence de l'autre alarme a eu lieu, notre méthode suggère de supprimer l'une des deux occurrences sans imposer laquelle supprimer. Dans notre cas, cela revient à supprimer les occurrences de a_1 qui précèdent une occurrence de a_2 (ou de a_5) et de supprimer une occurrence de a_2 ou de a_5 à chaque fois que a_2a_5 apparaît dans $Proj_{\{a_1,a_2,a_5\}}(S)$.

Etant donné que $Proj_{\{a_1,a_2,a_5\}}(S) = a_1a_1a_2a_5a_1a_2a_5a_1a_2a_5a_1a_2a_5a_1a_2a_5$, les séquences proposées $\widetilde{S_{7_i}}$ sont les suivantes :

- $\widetilde{S_{7_1}} = a_1\cancel{a_1}a_2\cancel{a_5}a_1a_2\cancel{a_5}a_1a_2\cancel{a_5}a_1a_2\cancel{a_5}a_1a_2\cancel{a_5} = a_1a_2a_2a_2a_2a_2$;
- $\widetilde{S_{7_2}} = a_1\cancel{a_1}\cancel{a_2}a_5\cancel{a_1}\cancel{a_2}a_5\cancel{a_1}\cancel{a_2}a_5\cancel{a_1}\cancel{a_2}a_5\cancel{a_1}\cancel{a_2}a_5 = a_1a_5a_5a_5a_5a_5$;
- $\widetilde{S_{7_3}} = a_1\cancel{a_1}\cancel{a_2}a_5\cancel{a_1}a_2\cancel{a_5}a_1a_2\cancel{a_5}a_1a_2\cancel{a_5}a_1a_2\cancel{a_5} = a_1a_5a_2a_2a_2a_2$;
- $\widetilde{S_{7_4}} = a_1\cancel{a_1}a_2\cancel{a_5}a_1\cancel{a_2}a_5\cancel{a_1}\cancel{a_2}a_5\cancel{a_1}a_2\cancel{a_5}a_1a_2\cancel{a_5} = a_1a_2a_5a_5a_2a_2$;
- …

De cette manière, il est possible de construire 32 séquences $\widetilde{S_{7_\iota}}$ différentes. En utilisant la connaissance dont il dispose sur les alarmes, l'expert n'aura pas à analyser les 32 séquences, il peut simplement définir une règle telle que « nous gardons la première occurrence qui apparaît » dans le cas d'une mutuelle dépendance.

En considérant la première séquence $\widetilde{S_{7_1}}$, nous avons $Proj_{\{a_1,a_2,a_5\}}\left(Proj_{\{a_1,a_2,a_5\}}(S)\right) = a_1a_1a_2a_5a_1a_2a_5\boldsymbol{a_1a_2a_5}a_1a_2a_5a_1a_2a_5$ et $Proj_{\{a_1,a_2,a_5\}}(\widetilde{S_{7_1}}) = a_1a_2a_2\boldsymbol{a_2}a_2a_2$, il existe donc bien un $\tilde{\sigma} = a_2$ tel que $\sigma = a_1a_2a_5$ puisse s'écrire de la forme $\sigma = a_1\tilde{\sigma}a_5$ avec $(a_1, a_5) \in (\{a_1, a_2, a_5\}^*)^2$. Nous avons donc bien $(Proj_{\Sigma_7}(S), \{f_7\}) \equiv (\widetilde{S_{7_1}}, \{f_7\})$. De plus, $|\widetilde{S_{7_1}}| < |Proj_{\Sigma_7}(S)|$.

$\widetilde{S_{7_1}}$ respecte donc les conditions de la définition des séquences équivalentes.

En effectuant cette démarche sur les 31 autres séquences, nous arrivons au résultat que tous les $\widetilde{S_{7_\iota}}$ respectent la définition des séquences équivalentes.

c. L'analyse experte

1. *Le cas des alarmes indépendantes*

Les alarmes a_4, a_7, a_8, a_9 et a_{10} ayant été jugées indépendantes des autres alarmes par notre méthode, l'expert ne les analyse pas. Comme ces alarmes ne sont pas analysées, pour assurer de conserver la sûreté du système, aucune de leur occurrence n'est supprimée dans le but d'obtenir une séquence équivalente permettant de réduire le flux d'alarmes.

2. *Le cas du fragment composé de a_3 et a_6*

L'alarme a_3 est l'alarme « Débit de combustible trop faible » et l'alarme a_6 est l'alarme « Vanne d'alimentation en combustible mal ouverte ». Une analyse experte du système permet de se rendre compte que lorsque la vanne d'alimentation n'est pas ouverte entièrement il n'y a systématiquement pas assez de carburant pour la combustion. Cela signifie que la connaissance que la vanne est mal ouverte est suffisante pour savoir que le débit est trop faible. Cependant, le niveau d'ouverture de la vanne étant calculé à travers une mesure du débit, le défaut de débit déclenche une alarme avant qu'une mauvaise ouverture de la vanne ne le fasse.

Une occurrence de a_6 implique donc que l'alarme a_3 est apparue auparavant. L'analyse experte valide donc la recommandation de la méthode qui consiste à supprimer les occurrences de a_3 qui précèdent une occurrence de a_6.

La séquence équivalente choisie par l'expert est donc celle proposée par notre méthode :

$$\widetilde{S_6} = \cancel{a_3}a_6\cancel{a_3}a_6a_3\cancel{a_3}a_6\cancel{a_3}a_6a_3 = a_6a_6a_3a_6a_6a_3$$

Le choix de cette séquence équivalente permet de supprimer 4 occurrences, et donc de diminuer la longueur de S de 4.

3. *Le cas du fragment composé de a_1, a_2 et a_5*

L'alarme a_1 est l'alarme « Pression haute dans la cuve de vapeur », l'alarme a_2 est « Pression très haute dans la cuve de vapeur », et l'alarme a_5 correspond à « Débit trop important en sortie de la cuve de vapeur ». Une analyse experte du système montre que le niveau de pression passe forcément par le seuil de déclenchement « haut » avant de passer par le seuil de déclenchement « très haut » car l'arme a_2 correspondant au seuil « très haut » est configurée de telle manière qu'elle disparaisse quand le niveau de pression passe en dessous du seuil « haut ».

Cette même analyse a également montré que le débit trop important en sortie de cuve est toujours dû à une pression ayant atteint le niveau « très haut », et que ces problèmes sont équivalents pour l'opérateur car ce dernier doit ouvrir la vanne de surpression au maximum dans tous les cas.

Ainsi, il est possible de conclure qu'une occurrence de a_2 implique qu'une occurrence de a_1 a eu lieu et que les alarmes a_2 et a_5 sont redondantes. Cette analyse permet donc de valider les recommandations de notre méthode qui consistent des actions suivantes :

- Supprimer les occurrences de a_1 qui précèdent une occurrence de a_2 ;
- Supprimer les occurrences de a_1 qui précèdent une occurrence de a_2 (ou de a_5) ;
- Supprimer une occurrence de a_2 ou de a_5 à chaque fois que a_2a_5 apparaît dans $Proj_{\{a_1,a_2,a_5\}}(S)$.

Ainsi, toutes les séquences équivalentes identifiées respectent bien l'analyse experte. Cependant, pour améliorer la réponse de l'opérateur, l'expert fait également un choix qui est de systématiquement garder la cause plutôt que la conséquence. C'est pourquoi l'expert choisit de supprimer les occurrences de a_5 au profit des occurrences de a_2. Ce choix correspond à la séquence équivalente $\widetilde{S_{7_1}}$:

$$\widetilde{S_{7_1}} = a_1\cancel{a_1}a_2\cancel{a_5}\cancel{a_1}a_2\cancel{a_5}\cancel{a_1}a_2\cancel{a_5}\cancel{a_1}a_2\cancel{a_5}\cancel{a_1}a_2\cancel{a_5} = a_1a_2a_2a_2a_2a_2$$

Le choix de cette séquence équivalente permet de supprimer 10 occurrences, et donc de diminuer la longueur de S de 10.

4. *La séquence équivalente finale*

Etant donné que nous cherchons à réduire la longueur de la séquence d'alarmes, nous considérons uniquement des suppressions d'occurrences lorsque nous cherchons à obtenir une séquence équivalente.

La séquence $\tilde{S}$ équivalente à S doit respecter la propriété suivante :

$$\forall\, \widetilde{S_\iota}\,, \begin{cases} Proj_{\Sigma_i}(\tilde{S}) = \widetilde{S_\iota} \\ \tilde{S}\ obtenue\ par\ suppression\ d'occurrences\ de\ S \end{cases}$$

Ainsi, les propriétés suivantes doivent être respectées :

- $Proj_{\{a_4\}}(\tilde{S}) = \widetilde{S_1} = a_4a_4a_4a_4$;
- $Proj_{\{a_7\}}(\tilde{S}) = \widetilde{S_2} = a_7a_7a_7a_7$;
- $Proj_{\{a_8\}}(\tilde{S}) = \widetilde{S_3} = a_8a_8a_8a_8$;
- $Proj_{\{a_9\}}(\tilde{S}) = \widetilde{S_4} = a_9a_9a_9a_9a_9$;
- $Proj_{\{a_{10}\}}(\tilde{S}) = \widetilde{S_5} = a_{10}a_{10}a_{10}a_{10}$;
- $Proj_{\{a_3,a_6\}}(\tilde{S}) = \widetilde{S_6} = a_6a_6a_3a_6a_6a_3$;
- $Proj_{\{a_1,a_2,a_5\}}(\tilde{S}) = \widetilde{S_{7_1}} = a_1a_2a_2a_2a_2a_2$;
- $\tilde{S}\ obtenu\ par\ suppression\ d'occurrences\ de\ S$;

La seule séquence qui respecte ces critères est la séquence $\tilde{S}$ ci-dessous :

$$\tilde{S} = a_1a_8a_9a_2a_6a_4a_2a_{10}a_7a_{10}a_4a_2a_7a_4a_9a_6a_9a_3a_4a_7a_8a_9a_2a_8a_8a_7a_6a_{10}a_{10}a_6a_9a_3a_2$$

Cette séquence est de longueur 33 occurrences, ce qui est **14 occurrences de moins que *S*** et représente une **réduction de 29,8% du flux d'alarmes**. Ces 14 occurrences correspondent bien à la somme des 4 occurrences supprimées dans l'analyse du fragment composé de a_3 et a_6, et des 10 occurrences supprimées dans l'analyse de fragment composé de a_1, a_2 et a_5.

5) Conclusions

La méthode développée dans cette partie a été présentée dans (Laumonier et al. 2019) et permet de réduire le flux d'alarmes arrivant aux opérateurs grâce à l'identification de toutes les précédences systématiques de type domination et mutuelle dépendance. Cette méthode trouve sa force dans le fait qu'elle identifie des relations entre des alarmes dont les occurrences ne sont pas adjacentes et qu'elle facilite grandement l'analyse de l'expert. En effet, en plus de guider l'expert en s'appuyant sur une représentation sous forme de réseaux de Petri, cette méthode diminue le nombre de séquences équivalentes parmi lesquelles l'expert va choisir la plus petite séquence qui conserve la sûreté du système.

L'inconvénient de cette méthode est que l'étape de nettoyage des places implicites est nécessaire pour avoir des fragments lisibles facilement, mais que la complexité polynomiale de cette étape entraîne un temps de calcul assez long.

Sur l'exemple de la chaudière à vapeur, l'utilisation de précédences systématiques semble plus efficace que l'utilisation des adjacences fréquentes. Cependant, pour comparer réellement ces deux approches nous allons, dans le chapitre suivant, les tester sur un jeu de données issu d'une installation industrielle réelle.

Chapitre 5 : Traitement d’une étude de cas industrielle et comparaison des méthodes

Dans ce chapitre, nous allons montrer comment les travaux développés durant cette thèse peuvent être utilisés dans l'industrie et comment réagissent les différents algorithmes sur des données industrielles. Pour ce faire, nous présentons un jeu de données qui sera analysé par les méthodes de recherche de séquences fréquentes et de recherche de précédences systématiques.

Grâce à une comparaison des résultats de ces deux méthodes, nous expliciterons la solution privilégiée pour les développements futurs ainsi que différentes améliorations possibles.

1) La gestion industrielle des historiques d'alarmes

a. Le stockage et l'utilisation des données

Les historiques d'alarmes sont des données temporelles où les occurrences d'alarmes sont enregistrées avec leur date d'apparition. Cette particularité permet de stocker ces données dans le service *Timeseries* que propose General Electric et qui est supporté par la plateforme Predix.

Le service *Timeseries* présente différents avantages qui facilitent l'utilisation des données qu'il stocke :

- Les données sont indexées pour pouvoir être rapidement retrouvées ;
- L'accès aux données est standardisé au travers d'API ;
- Les données sont disponibles depuis n'importe où par simple requête HTTP (avec l'autorisation adéquate).

Ces avantages ont facilité le traitement des différents historiques alarmes en permettant d'uniformiser le format des données d'entrée de l'algorithme. Ainsi, les codes développés sont génériques et indépendants de l'origine des données. Ces codes ont été écrits en Python 3.5 étant donné que ce langage est directement supporté par Predix.

Les collaborations avec plusieurs équipes internes à General Electric, et notamment la collaboration avec l'équipe *Brillant Manufacturing*, ont permis d'obtenir des jeux de données provenant de différentes industries (centrales électriques, usine minière et transport ferroviaire) afin de valider la généricité des algorithmes développés.

Parmi les différents jeux de données, pour proposer une étude comparative des deux méthodes développées durant cette thèse, nous détaillerons et utiliserons celui qui a mené aux résultats qui ont bénéficié de l'expertise la plus approfondie.

b. Les données utilisées pour l'étude comparative

Les données industrielles utilisées pour l'étude comparative proviennent d'une centrale électrique thermique fonctionnant selon un cycle combiné. Une centrale thermique à cycle combiné ressemble à une centrale thermique classique, telle que présentée dans le Chapitre 3, à la différence qu'elle associe deux types de turbines : une turbine à combustion et une turbine à vapeur. La turbine à combustion est actionnée par les gaz issus de la combustion et la turbine à vapeur est actionnée par la vapeur créée dans une chaudière grâce à la chaleur rémanente des fumées produites par la combustion.

Ce jeu de données est un historique d'alarmes représentant un mois de fonctionnement de la centrale durant lequel plusieurs avalanches d'alarmes ont eu lieu. Il est composé des 55.000 occurrences d'alarmes pour 489 alarmes différentes. L'évolution du nombre d'occurrences durant ce mois de fonctionnement (avec une période d'échantillonnage de 40 minutes) est représentée Figure 26 :

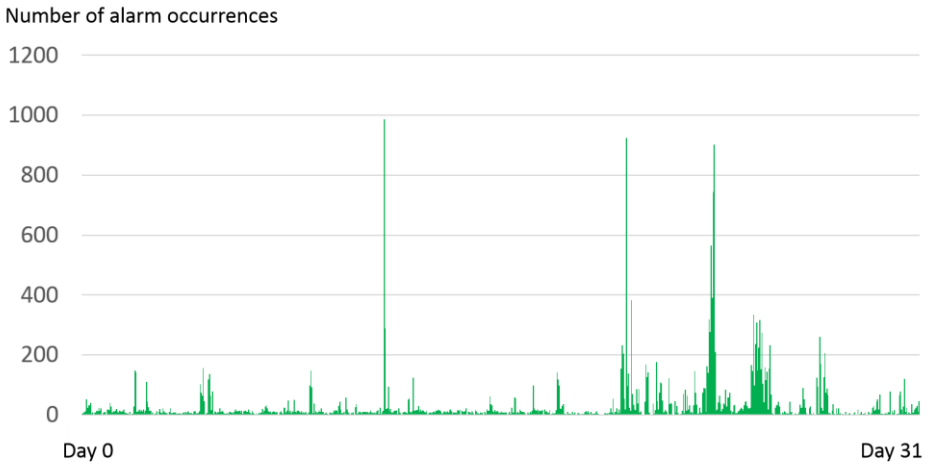


Figure 26 : Evolution du nombre d'occurrences d'alarmes durant le mois

La valeur moyenne du nombre d'occurrences d'alarmes par jour de cet historique est de 1.800 occurrences. Cette valeur est largement supérieure à la valeur maximale de 144 occurrences par jour recommandée par l'EEMUA. De même, alors que l'EEMUA recommande un maximum de 10 occurrences par 10 minutes, le jeu de données présente une valeur maximale de 460 occurrences en 10 minutes. Les recommandations de l'EEMUA sont donc largement dépassées et le besoin de réduire le flux d'alarmes pour les opérateurs est bien réel sur ce système.

Par la suite, ce jeu de données industriel est analysé à l'aide des deux méthodes développées durant cette thèse de manière à déterminer quelle méthode permet de réduire au mieux la charge des opérateurs.

2) Application de la méthode de recherche d'adjacences fréquentes

a. Application sur un jeu de données industriel

La première méthode appliquée pour identifier les relations existantes au sein de l'historique d'alarmes est la recherche d'adjacences fréquentes sous la forme de séquences fréquentes. Tel que détaillé dans le Chapitre 3, pour pouvoir appliquer cette méthode, la séquence d'alarmes représentant l'historique doit être découpée en différentes fenêtres d'observations.

La séquence d'alarmes étant relativement longue (55.000 occurrences), le découpage en fenêtres a peu de chance de systématiquement couper une séquence fréquente et de faire en sorte qu'elle ne soit pas identifiée par l'analyse.

Après avoir effectué différents tests, le compromis entre la vitesse de calcul et l'exhaustivité des résultats qui a été choisi est de prendre une fenêtre de taille 50 occurrences. Pour ne pas superposer les fenêtres d'observation, le pas choisi est également de 50 occurrences. Ainsi, la séquence d'alarmes est partitionnée en 1.100 fenêtres. La valeur maximale du support que peut avoir une séquence fréquente est donc de 1.100 (si la séquence en question est présente dans toutes les fenêtres d'observation).

En effectuant la recherche de séquences fréquentes selon l'algorithme *AprioriAll* présenté dans (Agrawal & Srikant 1995), 1.050 séquences fréquentes sont identifiées après 6 heures de calcul. Etant donné que cette analyse est faite hors ligne, ce temps de calcul est compatible avec notre contexte industriel.

La séquence fréquente la plus souvent observée correspond à une séquence de 2 occurrences d'alarmes. Cette séquence fréquente a été observée 763 fois. L'analyse experte a validé ce résultat et atteste de l'existence d'une relation d'équivalence (due à une mauvaise configuration) entre les deux alarmes qui composent cette séquence, il est donc possible de ne garder qu'une des deux alarmes à la place de la séquence. De ce fait, il serait possible d'économiser 763 occurrences d'alarmes sur le flux d'alarmes.

Sur les 1.050 séquences fréquentes, 80% des séquences ont un support inférieur à 10 et 60% de ces séquences ont un support inférieur à 5. Ces proportions indiquent qu'une grande partie des séquences identifiées ne présentent que peu ou pas d'intérêt pour la réduction du flux

d'alarmes : Economiser 5 occurrences sur un mois de fonctionnement (0,009% de la séquence d'alarmes totale) ne justifie pas forcément l'effort demandé pour l'analyse experte nécessaire à la validation de ces résultats.

Quelques séquences fréquentes identifiées sont représentées dans le Tableau 10 par leur support et leur longueur. De plus, ce tableau les ordonne en fonction de la valeur de leur support et indique la réduction possible sur le flux d'alarmes :

Tableau 10 : Séquences fréquentes identifiées

Valeur du support	Longueur de la séquence	Nombre d'occurrences filtrables
763	2	1*763
383	5	3*383
373	3	2*373
357	4	3*357
306	2	1*306
282	2	1*282
…	…	…

Les résultats indiqués par le Tableau 10, et notamment le nombre d'occurrences qui peuvent être retirées du flux d'alarmes, sont les résultats validés par un expert. Il est possible de remarquer que dans le cas de la séquence fréquente de longueur 5 et de support 383, l'expert a fait le choix de ne pas conserver qu'une seule alarme mais d'en conserver deux. Pour comprendre la décision de l'expert, nous détaillons les phénomènes qui se cachent derrière cette séquence :

Cette séquence est composée de 5 alarmes a_1, a_2, a_3, a_4 et a_5. En analysant la séquence, l'expert s'est rendu compte que a_1, a_2 et a_3 sont 3 alarmes en amont d'un processus de vote associé à trois mesures de trois capteurs de température redondants. a_4 est l'alarme en aval de ce vote et est configurée pour se déclencher quand au moins 2 des 3 alarmes en amont sont déclenchées. Lors de son analyse, l'expert a décidé que les alarmes a_1, a_2 et a_3 sont dues à une mauvaise configuration du système est que l'alarme a_4 est suffisante pour signaler un problème de température trop élevée. L'alarme a_5 quant à elle est une alarme de niveau hiérarchique supérieur qui peut avoir d'autres causes que a_4. De ce fait, l'expert a choisi de conserver l'alarme a_5 en plus de l'alarme a_4. Ces décisions permettent donc de réduire le flux d'alarmes de 3*383=1.149 occurrences.

La représentation du flux d'alarmes avant et après la réduction permise par les séquences fréquentes est montrée Figure 27 :

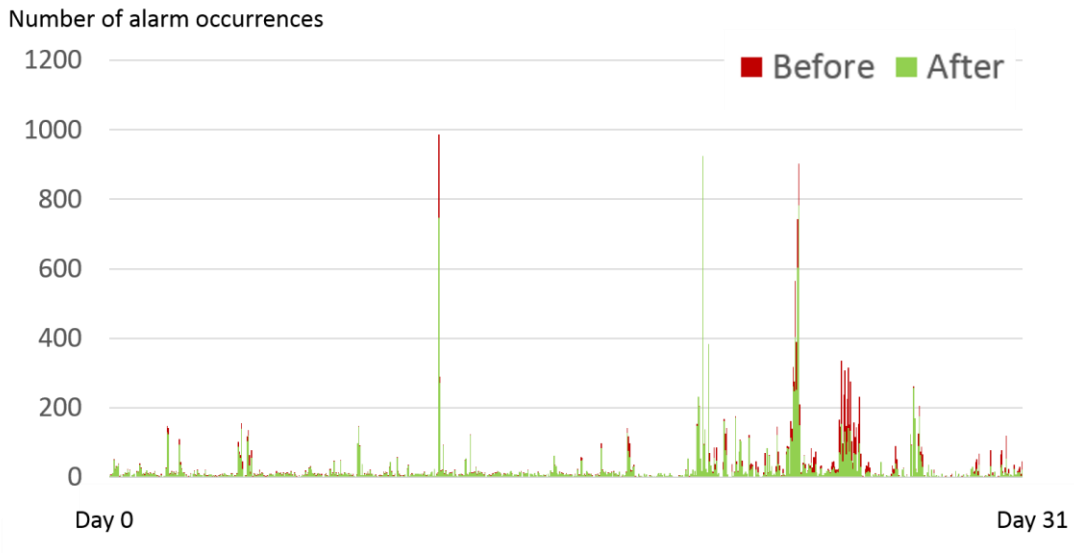


Figure 27 : Evolution du nombre d'occurrences d'alarmes durant le mois avant et après la réduction des séquences fréquentes

La Figure 27 superpose l'évolution du nombre d'occurrences observée d'après l'historique d'alarmes et l'évolution théorique où les occurrences sont supprimées selon l'analyse des séquences fréquentes. Ainsi, le gain que permet cette méthode sur le flux d'alarmes apparaît visuellement la figure. Nous constatons que cette méthode a permis une réduction de 6% sur la moyenne journalière, mais surtout de 35% lors des 10 minutes avec le plus d'occurrences, ce qui était bien l'objectif majeur de nos travaux.

Aux vues des réductions que permet cette approche, cette dernière rend plus fluide le travail des opérateurs en diminuant légèrement le nombre d'alarmes. Cependant, même avec cette méthode, le système d'alarmes reste loin de respecter les recommandations de l'EEMUA.

Un inconvénient de cette méthode est que l'expert doit effectuer une analyse longue et fastidieuse des différentes séquences fréquentes détectées. En effet, en plus du grand nombre de séquences identifiées, les alarmes et leurs occurrences peuvent intervenir dans différentes séquences fréquentes. De ce fait, les experts n'ont qu'une vision partielle du comportement des alarmes lorsqu'ils étudient une séquence fréquente particulière.

Par la suite, nous présentons l'implémentation qui a été réalisée sur la plateforme Predix de General Electric.

b. Implantation dans la plateforme Predix

La méthode consistant à rechercher des séquences fréquentes au sein d'une séquence d'alarmes a été implémentée sur la plateforme Predix. Cette implémentation a été réalisée dans le cadre d'une application servant de démonstration des capacités de la plateforme et des capacités techniques et scientifiques de General Electric. Elle est destinée aux clients actuels et futurs de General Electric.

L'allure globale de l'interface utilisateur de l'application est donnée sur la Figure 28. Cette application est une simple page de navigateur Internet qui va communiquer avec les serveurs Predix, notamment pour récupérer les données, lancer la recherche des séquences fréquentes et récupérer les résultats de l'analyse. Cette application est disponible depuis n'importe quel navigateur Internet sur lequel l'utilisateur a été identifié de façon sécurisée.

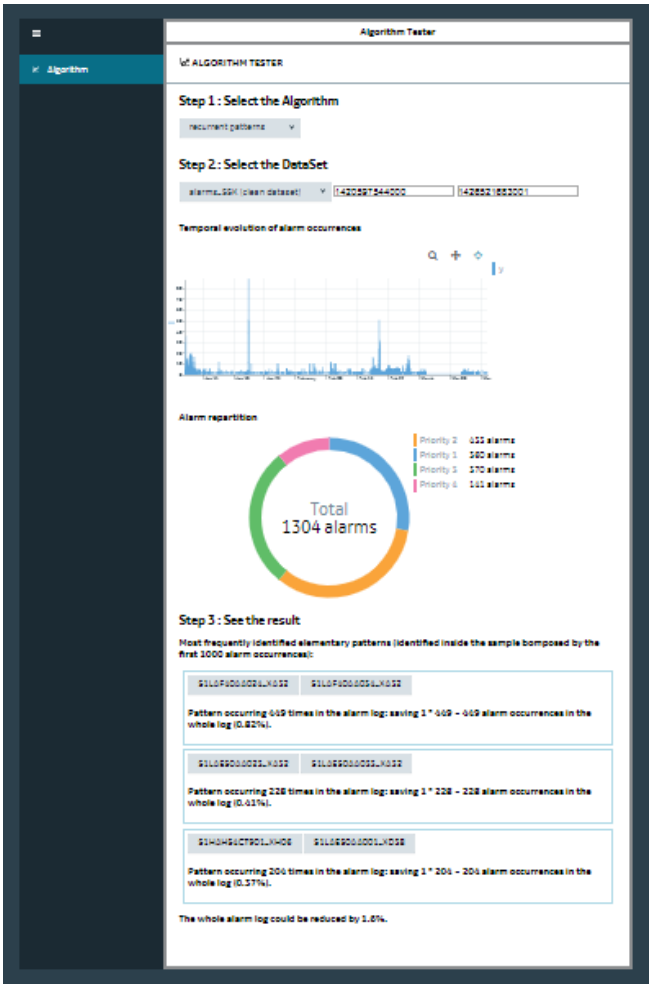


Figure 28 : Rendu visuel de l'application

Cette application se décompose en 4 phases. La première phase correspond au choix de l'algorithme et du jeu de données, et est affichée Figure 29:

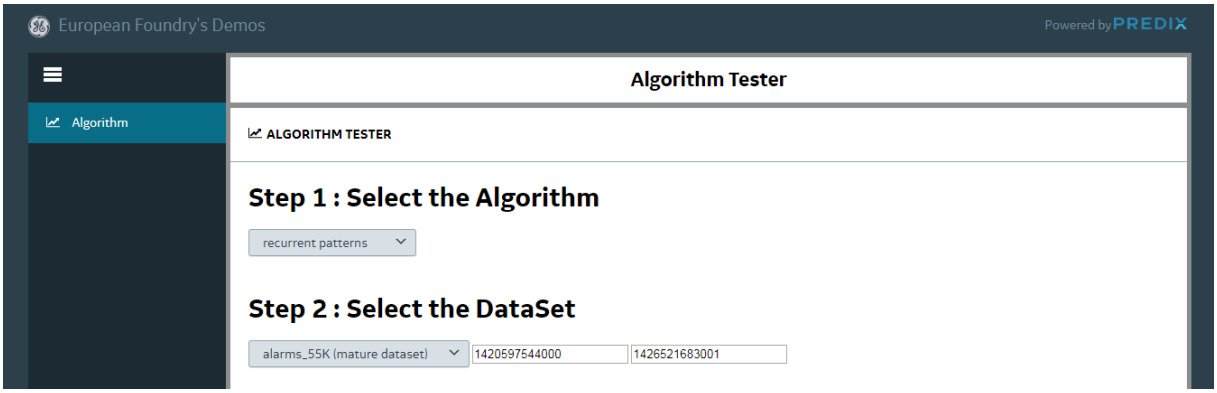


Figure 29 : Etapes d'initialisation de l'application

Cette partie correspond à la phase d'initialisation durant laquelle l'utilisateur va venir choisir l'algorithme à utiliser pour l'analyse et va également sélectionner le jeu de données qui va être étudié. L'algorithme de recherche de séquences fréquentes a été développé en Python 3.5 et est ici appelé « recurrent patterns ». Ce nom a été choisi car, pour une démonstration à un client, la notion de motifs récurrents est plus simple à appréhender que la notion de séquences fréquentes (qui nécessite de rentrer en détail dans les choix de modélisation). Cette application a été développée dans l'idée de regrouper différents algorithmes et outils d'analyse.

Le jeu de données présenté sur la Figure 29 est intitulé « alarms_55K (mature dataset) » et est un historique d'alarmes contenant 55.000 occurrences pour 1.304 alarmes différentes. Il représente approximativement deux mois de fonctionnement d'une centrale électrique thermique. Ce jeu de données est qualifié de mature car le système d'alarmes associé a déjà vécu plusieurs étapes de rationalisation et que de ce fait l'historique d'alarmes ne contient pas de phénomènes problématiques évidents tels que des alarmes bruyantes. Cela signifie également que, pour réduire le volume de cet historique, une analyse plus approfondie telle que la recherche de séquences fréquentes est nécessaire.

A côté du sélecteur du jeu de données sont indiquées la date de la première occurrence et la date de la dernière occurrence. Ces valeurs sont indiquées en temps Unix. En modifiant ces valeurs, il est possible de choisir la partie du jeu de données à étudier.

La phase suivante de l'application est présentée sur la Figure 30 et permet de visualiser l'évolution temporelle du nombre d'occurrences d'alarmes :

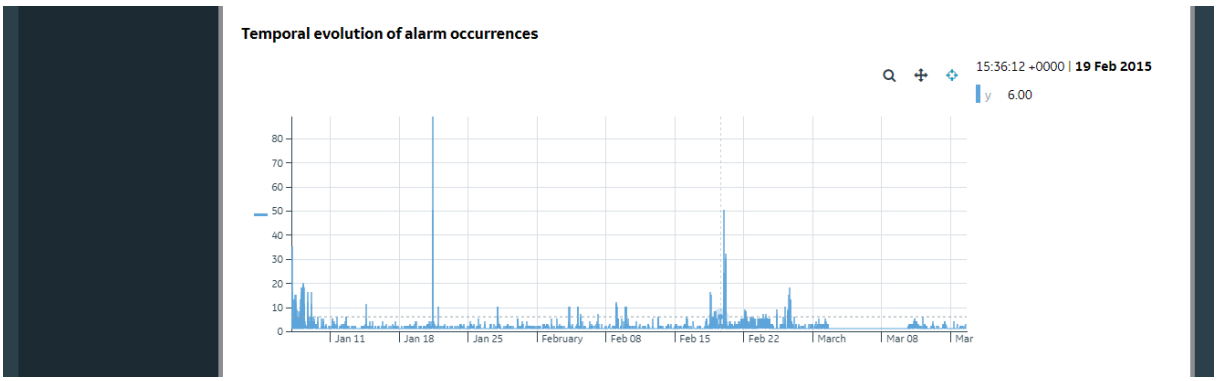


Figure 30 : Affichage de l'évolution temporelle du nombre d'occurrences d'alarmes

Cette visualisation permet d'identifier immédiatement les moments où un problème générant un nombre élevé d'alarmes a pu survenir. Ainsi, le pic présent aux alentours du 21 janvier pourrait représenter un changement de configuration partiel du système d'alarmes (le redémarrage après un changement de configuration déclenche les alarmes modifiées), et l'absence d'occurrence d'alarmes début mars peut indiquer un arrêt volontaire du système d'alarmes ou un arrêt du système d'acquisition. Il est également possible de visualiser les avalanches d'alarmes. Par exemple, le fait que lors des deux dernières semaines de février, il y ait eu beaucoup plus d'occurrences d'alarmes que le mois précédent peut indiquer la présence de plusieurs avalanches d'alarmes.

Pour pouvoir quantifier et identifier correctement le début des avalanches d'alarmes, le graphe est interactif et permet de déplacer le curseur de la souris pour afficher précisément le nombre d'occurrences à une date précise (indiqué en haut à droite sur la Figure 30).

Après cette étude qualitative des occurrences d'alarmes, la troisième phase permet d'effectuer une analyse rapide du système d'alarme et des alarmes configurées. Cette analyse est montrée sur la Figure 31 :

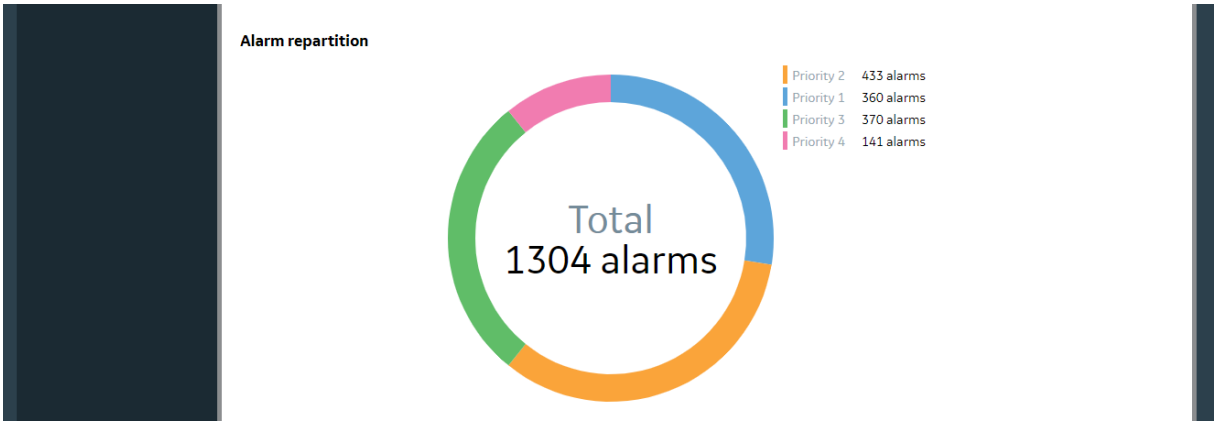


Figure 31 : Répartition des alarmes configurées en fonction de leur priorité

L'étude de la répartition des alarmes configurées en fonction de leur niveau de priorité permet de savoir en partie si le système d'alarmes respecte sa philosophie. En effet, dans le cas du jeu de données « alarms_55K (mature dataset) », 360 alarmes sur les 1.304 au total sont des alarmes de priorité 1. Ce niveau de priorité correspond au niveau le plus élevé, et les alarmes qui disposent de cette priorité doivent être des alarmes critiques pour le système. Le fait que plus d'un quart des alarmes configurées soit de priorité 1 indique un problème vis-à-vis de la compréhension ou du respect de la philosophie du système d'alarmes.

Cette représentation permet également de visualiser d'un coup d'œil le nombre de niveaux de priorité dont dispose le système d'alarmes.

Finalement, après les analyses préliminaires, la dernière phase correspondant à l'analyse approfondie est montrée sur la Figure 32. Cette analyse est la recherche de séquences fréquentes, mais est cependant limitée pour permettre une utilisation en ligne et montrer les possibilités de la plateforme Predix.

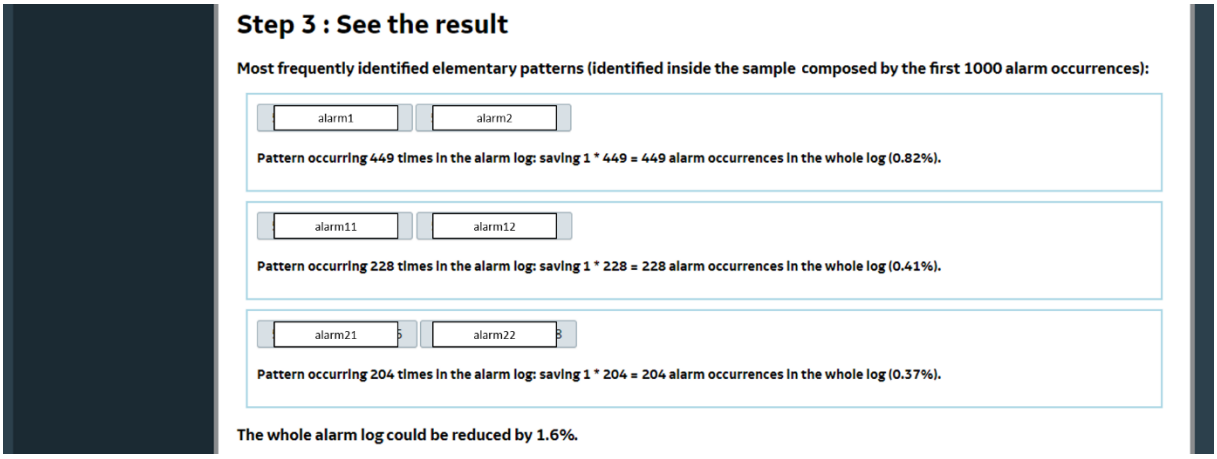


Figure 32 : Résultats de la recherche limitée de séquences fréquentes

Une démonstration devant être assez rapide, les résultats affichés ne sont pas les résultats que donnerait l'algorithme appliqué à l'entièreté de l'historique d'alarmes. Les résultats affichés sont obtenus en ne recherchant les séquences fréquentes que sur les 1.000 premières occurrences de l'historique, puis en extrapolant les résultats obtenus sur l'ensemble du jeu de données. L'application renvoie alors le résultat pour les trois séquences fréquentes avec la plus haute valeur de support. Par exemple, l'analyse sur les 1.000 premières occurrences a identifié la séquence fréquente *alarm1.alarm2*, puis la valeur du support de cette séquence (449) est déterminée indépendamment sur l'entièreté de l'historique. Enfin, sous l'hypothèse que tous les motifs sont préalablement validés par un expert, le gain sur la longueur totale de l'historique est déterminé. Ainsi, l'analyse de la seule séquence fréquente *alarm1.alarm2* permet

une réduction de 449/55.000 = 0,82 % du flux d'alarmes. Toujours dans le cadre de la démonstration, la réduction totale permise par les trois séquences fréquentes est calculée. Ainsi, avec cette analyse limitée permettrait de réduire de 1,6% la longueur de la séquence.

Une autre fonctionnalité qui a été développée, pour faciliter l'expertise, consiste dans la possibilité de pouvoir cliquer sur le nom des alarmes composant les motifs pour afficher des informations complémentaires. Ces informations peuvent être le niveau de priorité de l'alarme, le lieu de l'alarme dans le système ou toute autre information qui pourrait être utile pour l'analyse.

Cette démonstration a été réalisée à différentes personnes internes et externes à General Electric. Les retours ont permis de valider et de réaffirmer l'intérêt que présentent ces travaux pour résoudre le problème de la surabondance d'alarmes que doivent gérer les opérateurs.

3) Application de la méthode de recherche de précédences systématiques

a. Application sur un jeu de données industriel

La seconde méthode utilisée pour identifier les relations entre les alarmes de l'historique d'alarmes est la recherche de précédences systématiques par le biais de la recherche de dominations et de mutuelles dépendances. Comme il est décrit dans le Chapitre 4, les relations de dominations et de mutuelles dépendances trouvées sont ensuite traduites sous la forme d'un Réseau de Petri pour faciliter et guider l'analyse de l'expert.

La recherche des dominations et des mutuelles dépendances dans la séquence d'alarmes, ainsi que la construction du RdP traduisant ces relations, a été réalisée en 1 heure et 45 minutes. Comme pour la méthode précédente, ce temps de calcul est suffisamment court pour permettre son utilisation dans un contexte industriel. Après le calcul, nous avons identifié 36 fragments de RdP connexes qui représentent l'existence de relations entre 141 alarmes. 348 alarmes ont également été identifiées comme étant indépendantes, ces alarmes sont représentées par des transitions qui sont à la fois source et puits.

Ce sont les 36 fragments qui intéressent particulièrement l'expert pour pouvoir réduire le nombre et le flux d'alarmes. Pour effectuer son analyse, l'expert s'intéresse d'abord aux fragments dont les alarmes apparaissent le plus souvent au sein de la séquence d'alarmes. Ainsi, il commencera par le fragment a) sur la Figure 33 qui relie deux alarmes apparaissant 1.468 fois (a_1) et 1.051 fois (a_2), et finira par le fragment b) sur la Figure 33 qui relie 4 alarmes qui apparaissent toutes 11 fois (a_{138}, a_{139}, a_{140} et a_{141}).

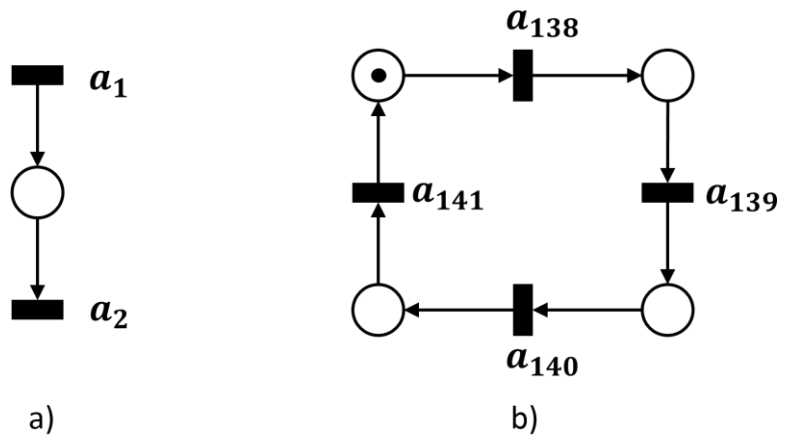


Figure 33 : Exemples de fragments obtenus a) avec alarmes occurrant le plus souvent, b) avec alarmes occurrant le moins souvent

Les deux types de fragments représentés sur la Figure 33 sont caractéristiques de la majorité des fragments identifiés : le fragment a) correspond à une domination et le fragment b) correspond à des mutuelles dépendances. Des fragments plus complexes qui relient des dominations et des mutuelles dépendances existent mais sont moins nombreux (5 fragments sur les 36 identifiés). Un exemple de fragment plus complexe est donné Figure 34 :

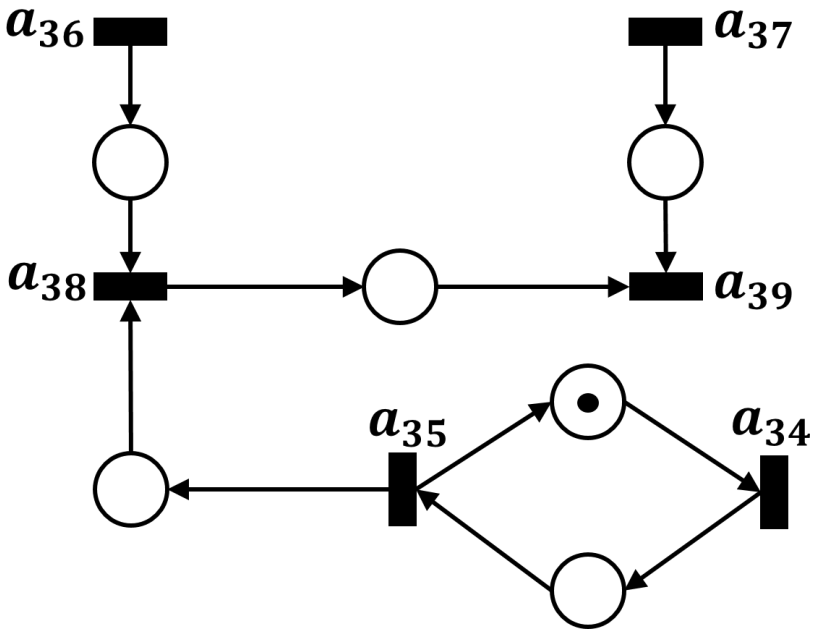


Figure 34 : Exemple de fragment composé de dominations et d'une mutuelle dépendance

Avant de détailler la composition du fragment présenté sur la Figure 34, nous rappelons que pour faciliter la lecture par l'expert, les places implicites (dont la suppression ne change rien à l'évolution du RdP) sont supprimées. De ce fait, certaines dominations et mutuelles

dépendances n'apparaissent pas sur ce fragment. Le fragment de la Figure 34 est composé des dominations et mutuelles dépendances suivantes :

- $a_{34} \leftrightarrows a_{35}$;
- $a_{34} \rightarrow a_{38}$ (place enlevée par réduction des places implicites) ;
- $a_{35} \rightarrow a_{38}$;
- $a_{34} \rightarrow a_{39}$ (place enlevée par réduction des places implicites) ;
- $a_{35} \rightarrow a_{39}$ (place enlevée par réduction des places implicites) ;
- $a_{36} \rightarrow a_{38}$;
- $a_{36} \rightarrow a_{39}$ (place enlevée par réduction des places implicites) ;
- $a_{38} \rightarrow a_{39}$;
- $a_{37} \rightarrow a_{39}$;

Pour comprendre ce fragment, nous allons détailler l'analyse experte qui a été effectuée sur les alarmes qui le composent. Tout d'abord, a_{34} est une alarme « température basse » associée à un seuil de température. Cette alarme se déclenche quand la valeur absolue de la température passe en dessous d'une certaine valeur T_1. a_{35} correspond à la même alarme « température basse » mais est associée au dépassement de la valeur relative de T_1, appelée $T_1^{\%} = \frac{T_1 * 100}{T_{max}}$ (T_{max} étant la valeur maximale mesurable par le capteur). La mutuelle dépendance $a_{34} \leftrightarrows a_{35}$ est donc parfaitement logique puisque que a_{34} et a_{35} sont redondantes. L'expert recommande de ne garder que a_{34}.

Ensuite, la chaîne de domination $a_{35} \rightarrow a_{38}$ et $a_{38} \rightarrow a_{39}$ correspond au fait que a_{38} est l'alarme « température très basse » et que a_{39} est l'alarme « température très très basse ». Lors du fonctionnement du processus, pour déclencher l'alarme « température très basse » il a forcément fallu déclencher auparavant l'alarme « température basse », et pour déclencher l'alarme « température très très basse » il a nécessairement fallu déclencher les alarmes « température basse » et « température très basse ». L'expert a attesté que cette conception est une erreur et qu'il est possible de filtrer a_{34}/a_{35} quand a_{38} est présent, et qu'il est possible de filtrer a_{34}/a_{35} et a_{38} quand a_{39} est présent.

Les dominations $a_{36} \rightarrow a_{38}$ et $a_{37} \rightarrow a_{39}$ s'expliquent par le fait que a_{36} et a_{37} sont des alarmes indiquant une panne électrique sur les vannes d'isolement entre les refroidisseurs (a_{36} associé à la vanne 1 et a_{37} associé à la vanne 2), et que la vanne 2 est celle qui est censée protéger le système si la vanne 1 défaille. Ainsi, si la vanne 1 ne fonctionne plus durant un laps de temps suffisant la température chute est l'alarme a_{38} se déclenche, et si les deux vannes ne fonctionnent pas suffisamment longtemps la température chute encore plus bas et a_{39} se

déclenche. L'expert a proposé de ne garder que les alarmes a_{36} et a_{37} en argumentant que l'action corrective est la même entre a_{38} et a_{36}, et entre a_{39} et (a_{36}, a_{37}).

La représentation du flux d'alarmes avant et après la réduction permise par l'étude approfondie des différents fragments est donnée Figure 35 :

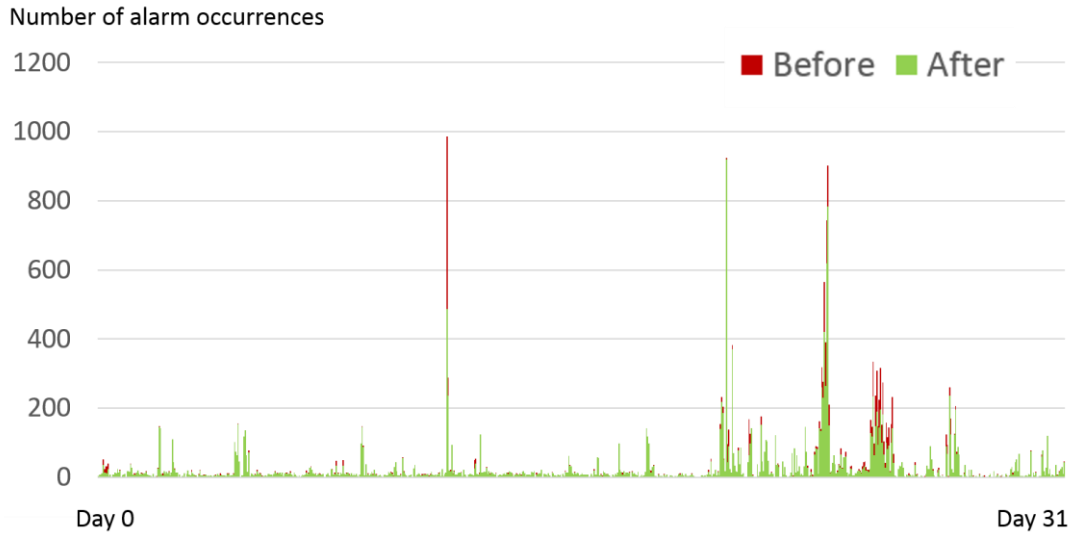


Figure 35 : Evolution du nombre d'occurrences d'alarmes durant le mois avant et après la réduction permise par les fragments de RdP

Comme dans la partie précédente, la Figure 35 superpose l'évolution du nombre d'occurrences observée d'après l'historique d'alarmes et l'évolution théorique où des occurrences sont supprimées d'après l'analyse des fragments représentant les dominations et les mutuelles dépendances. De cette manière, la réduction que permet cette méthode sur le flux d'alarmes apparaît visuellement le graphe. Sur ce jeu de données, cette méthode a permis une réduction de 5% sur la moyenne journalière et de 38% lors des 10 minutes comportant le plus d'occurrences.

Numériquement, les résultats obtenus sont similaires à ceux obtenus avec l'approche par recherche de séquences fréquentes. Pour pouvoir justifier du choix d'une méthode par rapport à l'autre, il est donc nécessaire de faire une étude comparative détaillée des deux approches à travers leurs résultats.

Avant d'effectuer cette étude comparative, nous détaillons comment l'implémentation industrielle de la recherche de précédences systématiques est réalisée.

b. Implantation dans la plateforme Predix

La deuxième méthode, consistant à chercher des dominations et des mutuelles dépendances dans un historique d'alarmes, est en cours d'implémentation. L'algorithme est codé en Python 3.5 et repose toujours sur la communication avec des données stockées dans *Timeseries*.

La plateforme Predix ayant évolué entre le moment où cette nouvelle méthode a été testée et le moment où la démonstration précédente a été créée, le choix a été fait d'implémenter cette méthode dans *Predix Asset Performance Management* (APM). Predix APM est une suite logicielle de services conçue pour optimiser la performance des équipements. C'est une solution générique qui s'adapte à toutes les industries et tous les équipements. Comme son nom l'indique, Predix APM est centré autour de la notion d'équipements et les données stockées dans un service *Timeseries* sont donc liées à des équipements : les équipements pouvant être de n'importe quelle taille allant d'une machine particulière à une usine.

Dans le cadre de l'implémentation de la deuxième méthode dans Predix APM, les données correspondant aux historiques d'alarmes ont été liées à des équipements et sont ainsi disponibles pour effectuer une analyse. Cependant, l'architecture de Predix APM étant différente de la plateforme Predix classique et étant donné que Predix APM ne dispose pas encore d'interface graphique pour afficher des réseaux de Petri, une discussion est en cours pour décider du format des résultats qui seront affichés dans le cadre d'une démonstration. De ce fait, l'implémentation de la méthode de recherche de précédences systématiques n'est pas encore finalisée.

4) Comparaison des deux méthodes

Après avoir observé individuellement les résultats de chacune des méthodes, il est nécessaire de les comparer pour faire ressortir leurs différences et leurs complémentarités.

a. Comparaison numérique des résultats

Pour évaluer l'intérêt d'une méthode par rapport à l'autre, nous nous intéressons d'abord à la comparaison des résultats en termes de nombre de motifs fréquents/fragments, de temps de calcul et de réduction du flux d'alarmes.

Pour ce jeu de données particuliers, la recherche de séquences fréquentes a duré 6 heures et a identifié 1.050 séquences fréquentes et la recherche de précédences systématiques a pris 1 heure et 45 minutes, et a retourné 36 fragments de RdP. La première méthode est donc plus

consommatrice en temps de calcul. Concernant la conclusion sur le nombre de séquences fréquentes comparativement au nombre de fragments, il faut d'abord se poser la question sur la manière dont ces résultats peuvent être comparés. Il y a donc besoin de faire une analyse plus fine sur des critères tels que le niveau de certitude et la complétude de ces séquences et fragments.

Pour comparer la réduction du flux d'alarmes permise par chacune de nos deux méthodes, la Figure 36 met en parallèle les résultats globaux de celles-ci :

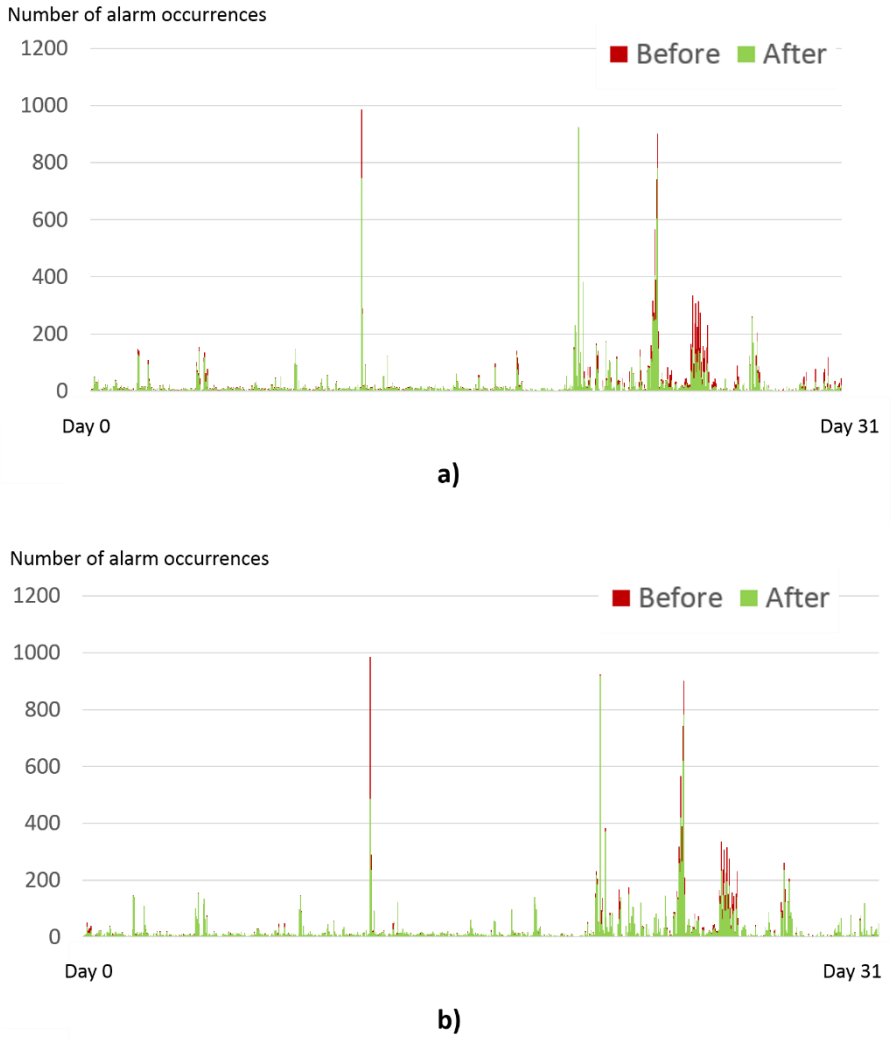


Figure 36 : Mise en parallèle des résultats des méthodes par recherche a) d'adjacences fréquentes et b) de précédences systématiques

Une analyse rapide de la Figure 36 permet de confirmer la similarité entre les résultats des deux méthodes. En effet, comme indiqué précédemment, la recherche de séquences fréquentes permet une réduction moyenne du flux d'alarmes de 6% et une réduction de 35% lors des 10

minutes les plus denses, et la recherche de précédences systématiques permet une réduction moyenne de 5% et une réduction de 38% lors des 10 minutes les plus denses.

Cependant, en regardant plus en détail cette figure, il est possible de constater que la réduction du flux d'alarmes ne se fait pas sur les mêmes parties de la séquence. Une analyse plus fine des résultats données par ces deux méthodes est donc nécessaire.

b. Comparaison qualitative et experte des résultats

En comparant les séquences fréquentes obtenues par la première méthode et les fragments de RdP obtenus par la seconde, certaines différences concernant les relations identifiées ressurgissent. Ces différences sont inhérentes à la nature des phénomènes recherchés par chacune des méthodes.

1. Les relations identifiées par l'une des deux méthodes seulement

La recherche de séquences fréquentes a pour but d'identifier des relations fréquentes mais pas forcément systématiques entre occurrences d'alarmes. Ainsi, si localement le système d'acquisition a connu un problème et que l'ordre des occurrences a été bouleversé, les relations peuvent être identifiées car ce problème n'impactera que la valeur du support. Cependant, cette méthode ne permet pas de trouver des relations entre alarmes dont les occurrences ne sont pas adjacentes car la recherche se fait sur des séquences et donc sur des alarmes adjacentes.

Au contraire, la recherche de précédences systématiques est réalisée sur des projections et peut donc relier des occurrences qui ne sont pas adjacentes. Cependant, si l'ordre des occurrences est modifié localement, il est possible de passer à côté de relations entre occurrences.

Les deux méthodes ne cherchent donc pas les mêmes phénomènes et sont complémentaires.

Pour faciliter la compréhension, nous illustrons ces propos par l'exemple développé dans les chapitres précédents :

Exemple :

$$S = a_1a_3a_1a_8a_9a_2a_5a_6a_4a_3a_1a_2a_5a_{10}a_7a_{10}a_4a_1a_2a_5a_7a_4a_9$$

$$a_6a_9a_3a_4a_7a_8a_1a_9a_2a_5a_8a_8a_7a_3a_6a_{10}a_3a_{10}a_6a_9a_1a_3a_2a_5$$

- La recherche de séquences fréquentes a identifié les séquences a_2a_5, a_3a_1, $a_1a_2a_5$ et $a_9a_2a_5$.

- La recherche de précédences systématiques a identifié le réseau de Petri rappelé sur la Figure 37.

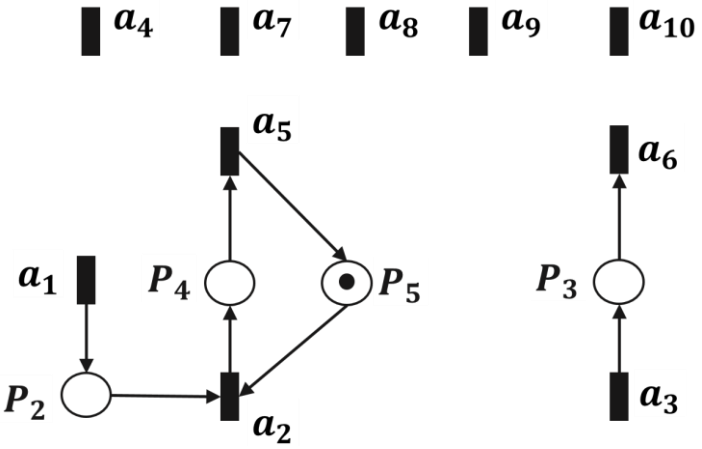


Figure 37 : Réseau de Petri identifié dans l'exemple de la chaudière

En comparant les résultats, il est possible de remarquer que la seconde méthode a trouvé des relations entre les occurrences de a_3 et de a_6 alors que cette relation est restée invisible pour la première méthode car a_3 et a_6 ne sont pas adjacentes au moins deux fois.

A l'opposé, à travers la séquence fréquente $a_9a_2a_5$, la première méthode a identifié une relation entre les occurrences de a_9 et les occurrences de a_2 et de a_5 alors que la deuxième méthode a jugé que les occurrences de a_9 étaient indépendantes des occurrences d'autres alarmes.

C'est au travers d'une expertise qu'il est possible de valider les résultats de l'une ou de l'autre de ces méthodes, et de conclure sur l'existence de relations entre les alarmes. Dans certains cas, des séquences fréquentes identifiées ne correspondront pas à des relations entre alarmes, et dans d'autres cas, ce seront les fragments qui ne représenteront pas complètement les relations entre alarmes. La raison pour laquelle les résultats des deux méthodes ne s'accordent pas toujours et que, en plus de ne pas identifier les mêmes types de relations, les méthodes ne présentent pas les mêmes niveaux de certitude et de complétude sur les résultats.

2. *Le niveau de certitude des résultats*

Il est important de garder en tête que la recherche de séquences fréquentes et la recherche de précédences systématiques ne sont pas suffisantes pour conclure et qu'une analyse experte des relations identifiées est nécessaire. Le critère de comparaison discuté dans cette section est le niveau de certitude avant l'expertise.

La méthode de recherche de séquences fréquentes identifiera une séquence si elle apparaît au moins deux fois et l'enregistrera comme séquence fréquente dans les résultats. De ce fait, les résultats de cette méthode ont potentiellement un faible niveau de certitude.

Il est possible d'illustrer ce propos à travers un exemple trouvé lors de l'analyse du jeu de données industriel :

Exemple :

Les observations suivantes ont pu être réalisées sur deux alarmes *alarm8* et *alarm15* :

- Une séquence fréquente *alarm8.alarm15* qui possède un support de 150 a été identifié par la première méthode.
- La seconde méthode a identifié *alarm8* et *alarm15* comme étant des alarmes indépendantes.
- Dans la séquence d'alarmes totale, *alarm8* est apparu 1.431 fois et *alarm2* est apparu 1.268 fois.

La valeur du support de la séquence fréquente est inférieure au nombre d'occurrences des alarmes qui la composent avec un facteur proche de 10. Il est donc très peu certain qu'au moment de l'expertise la séquence fréquente représente une relation réelle entre les deux alarmes.

Du fait de la manière dont les méthodes fonctionnent, les résultats de la recherche de précédences systématiques ont intrinsèquement un niveau de certitude plus élevé que ceux obtenus par la recherche de séquences fréquentes.

3. *La complétude des résultats*

Etant donné que la méthode centrée sur la recherche de précédences systématiques construit les RdPs en créant une bijection entre les alarmes et les transitions, l'unicité des alarmes est assurée au sein du RdP. De ce fait, un fragment trouvé par cette méthode regroupe l'ensemble des relations systématiques qui lient les alarmes de ce fragment. Les résultats de la seconde méthode sont donc complets.

Au contraire, la première méthode retourne des séquences fréquentes et, comme nous avons pu le voir au travers des exemples précédents, une alarme peut faire partie de plusieurs séquences fréquentes différentes. De ce fait, pour analyser une alarme et ses relations avec les autres alarmes, il faut analyser chaque motif qui contient cette alarme. Une séquence fréquente seule ne suffisant pas à effectuer l'analyse d'une alarme, un résultat sous la forme d'une séquence fréquente n'est donc pas complet.

Exemple :

En reprenant l'exemple de la chaudière rappelé précedemment, sous supposons que l'expert veut étudier l'alarme a_2 et ses potentielles relations avec d'autres alarmes :

- Avec la première méthode, l'expert doit analyser toutes les séquences fréquentes qui contiennent a_2 : a_2a_5, $a_1a_2a_5$ et $a_9a_2a_5$.
- Avec la deuxième méthode, l'expert doit analyser un seul fragment de RdP qui lie a_2 à toutes les alarmes pouvant être en relation avec a_2.

Ainsi, la méthode par recherche de précédences systématiques permet d'identifier des relations complexes qui s'étalent potentiellement sur un grand laps de temps. Ces relations, où des alarmes parasites ont pu apparaître entre les alarmes liées, sont représentées en un seul fragment alors que, dans le meilleur des cas, il faudrait analyser plusieurs séquences fréquentes pour les retrouver avec notre première méthode.

Notre seconde méthode est donc bien meilleure que celle par recherche de séquences fréquentes quand il s'agit de fournir des résultats complets. C'est en partie pour cette raison qu'il y a un grand écart entre le nombre de séquences fréquentes trouvées (1.050) et le nombre de fragments identifiés (55).

c. Résultats de la comparaison

La différence majeure entre les méthodes de recherche de séquences fréquentes et de précédences systématiques réside dans la facilité qu'a l'expert pour effectuer son analyse des résultats. A cela s'ajoute le fait que, lors de l'analyse des fragments issus de la recherche des précédences systématiques, l'expert est guidé dans son analyse par la représentation graphique. Cette différence est essentielle lorsqu'il est envisagé d'industrialiser et de commercialiser ces méthodes.

A part la différence concernant la facilité de l'expertise et la complétude des résultats, les deux méthodes permettent une réduction équivalente du flux d'alarmes. Mais bien que le résultat des deux méthodes soit comparable, ces méthodes cherchent à identifier des phénomènes différents et sont de ce fait complémentaires.

5) Conclusions et discussion

Après avoir confronté les méthodes de recherche de séquences fréquentes et recherche de précédences systématiques à un jeu de données industriel, et après avoir analysé les résultats obtenus avec un expert, nous pouvons conclure que les deux méthodes fonctionnent et trouvent leur véritable valeur dans la fluidification du travail de l'opérateur en permettant de réduire significativement le flux d'alarmes lors des avalanches. De plus, le grand volume de données utilisées a permis de constater que les algorithmes supportent bien le passage à l'échelle.

Les deux méthodes ont beau permettre des réductions équivalentes du flux d'alarmes, elles sont cependant différentes pour l'expert chargé d'analyser les résultats : permettant de visualiser directement les relations entre alarmes et offrant une analyse guidée beaucoup plus simple, la méthode par recherche de précédences systématiques est préférée. C'est pour cette raison qu'un brevet concernant l'industrialisation de cette méthode a été déposé en Europe et aux Etats-Unis.

Pour aller plus loin avec cette méthode, les experts ont également émis le souhait de voir apparaître sur les fragments de réseau de Petri des informations supplémentaires utiles à leur analyse. Un exemple de fragment avec les informations complémentaires demandées par des experts est présenté sur la Figure 38 :

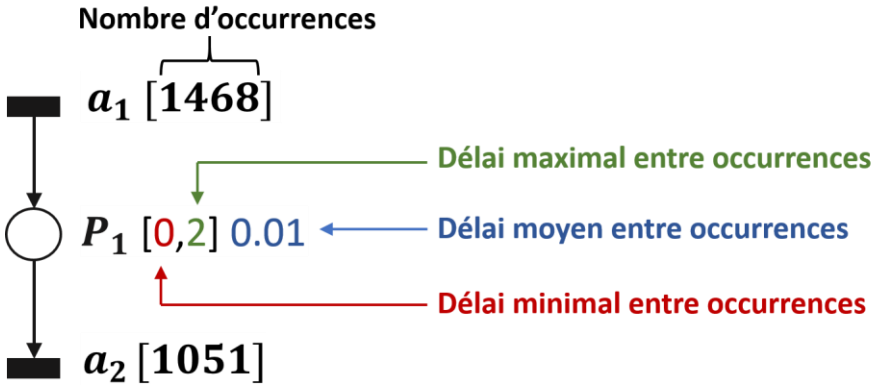


Figure 38 : Fragment avec informations utiles à l'expertise

Les informations que les experts ont voulu voir apparaître en priorité sont :

- Le nombre d'occurrences de chaque alarme : Pour qu'ils puissent définir l'ordre dans lequel ils vont analyser les fragments (plus une alarme apparaît souvent, plus la réduction potentielle sera importante).
- Les informations permettant de comprendre le comportement temporel existant entre les alarmes : Les informations retenues sont le délai minimal, le délai maximal et le

délai moyen (en secondes) entre une occurrence de l'alarme en amont et une occurrence de l'alarme en aval. La prise en compte du temps physique pour l'analyse de la séquence d'alarmes fait aussi partie des perspectives de ces travaux

Un phénomène qu'il faut garder en tête est que la qualité des résultats dépend grandement de la qualité du jeu de données. Par exemple, lorsque des tests ont été réalisés sur un jeu de données présentant des alarmes bruyantes, ces alarmes se sont retrouvées dans une majorité des séquences fréquentes et ces alarmes ont également créé un grand nombre de relations de dominations (liant ainsi des fragments normalement indépendants). Dans ce cas, il serait donc nécessaire de filtrer les alarmes bruyantes avant d'appliquer nos méthodes.

De même, lorsque des tests ont été effectués sur des données provenant de simulateur, les résultats montraient une majorité de grands fragments contenant de dix à trente alarmes. Ce phénomène n'avait pas été constaté avec les données provenant de l'observation du processus réel. L'analyse de l'expert a validé les doutes en indiquant que les alarmes étaient bien liées par le processus mais qu'elles ne menaient pas à une réduction du flux d'alarmes. En effet, ces fragments représentent des scénarios de défauts avec des séquences systématiquement générées par le simulateur. Elles ne prennent donc pas en compte la part d'aléatoire qui existe lors de l'évolution d'un processus industriel réel.

Il est donc important de prévoir les conséquences qu'aura un jeu de données mal adapté avant de lancer les calculs.

Conclusions générales et perspectives

1) Résumé des travaux

L'objectif de cette thèse est de proposer des contributions pour résoudre la problématique du trop grand nombre d'alarmes industrielles qui doivent être traitées par les opérateurs. Pour ce faire, nous avons limité nos travaux à l'utilisation d'historiques d'alarmes pour découvrir des relations entre les alarmes. En modélisant les apparitions d'alarmes comme des évènements dont les occurrences forment une séquence d'alarmes, nous avons recherché deux types de relations :

- Les adjacences fréquentes : En supposant que les alarmes qui transmettent des informations redondantes pour l'opérateur apparaissent consécutivement, nous avons cherché toutes les séquences d'occurrences adjacentes qui se produisent plusieurs fois dans la séquence d'alarmes.
- Les précédences systématiques : En supposant que si une relation entre deux alarmes existe, alors elle ne dépend pas des autres alarmes, nous avons cherché à vérifier pour toutes les alarmes s'il existait une alarme dont l'occurrence était une condition nécessaire à l'apparition d'une autre alarme.

Pour réduire le flux d'alarmes, nous proposons de filtrer les occurrences d'alarmes qui ne seraient pas utiles à l'opérateur. Cependant, le problème de la réduction du flux d'alarmes ne peut pas être découplé du problème de la sûreté de fonctionnement. C'est pourquoi, dans l'approche que nous avons développée, c'est un expert du processus industriel utilisant l'historique d'alarmes qui décide des occurrences à filtrer après avoir analysé différentes propositions issues de nos méthodes formalisées.

Pour pouvoir valider notre approche, nous l'avons confrontée à différents historiques d'alarmes obtenus par le biais de la collaboration industrielle avec General Electric. Cette collaboration a également été utile pour avoir accès à l'expertise industrielle et aller jusqu'à obtenir la valeur de la réduction permise sur les séquences d'alarmes.

Ces travaux ont fait l'objet d'une implantation sur la plateforme Predix de General Electric et les premiers résultats ont été si encourageants qu'un dépôt de brevet en Europe et aux Etats-Unis a été initié. Cependant, la réduction que permet notre approche sur le flux d'alarmes n'est pas complètement suffisante et des efforts supplémentaires doivent encore être réalisés pour respecter les recommandations de (EEMUA 2007).

2) Conclusions

Les deux méthodes formalisées développées durant cette thèse permettent des réductions similaires du flux d'alarmes : une réduction moyenne du flux d'alarmes d'environ 5% et une réduction d'environ 35% lors des 10 minutes les plus denses. Comme ces résultats l'indiquent, la vraie valeur ajoutée de nos méthodes n'est pas dans la diminution globale du flux d'alarmes, mais dans la fluidification du travail des opérateurs en réduisant efficacement le nombre d'alarmes lors des avalanches.

De plus, bien que les deux méthodes développées identifient pour une part les mêmes relations, chacune des méthodes permet d'obtenir des relations qui échappaient à l'autre. Nos méthodes sont donc partiellement complémentaires.

La grosse différence entre ces méthodes se situe dans la qualité de l'analyse experte : dans un cas, un grand volume de relations statistiques et de potentielles solutions pour réduire la séquence d'alarmes est donné à l'expert pour analyse, alors que dans l'autre cas, des règles précises pour réduire la séquence d'alarmes et un réseau de Petri traduisant les relations systématiques lui sont confiés.

Ainsi la deuxième méthode qui consiste à identifier les précédences systématiques est beaucoup plus utile pour l'expert qui voit son analyse guidée par le formalisme de la méthode. Cette méthode présente également l'avantage de pouvoir évoluer pour prendre en compte de nouveaux phénomènes tels que le temps ou la connaissance de l'évolution du processus.

3) Perspectives

La suite logique de ces travaux est d'implémenter notre approche dans un système d'alarmes industriel pour vérifier la viabilité de notre approche dans un environnement réel.

En outre, ces travaux ont été développés avec pour contrainte de ne connaître que la séquence d'alarmes. Or, d'autres informations sont disponibles aussi bien au sein de l'historique d'alarmes qu'autour du système d'alarmes.

a. L'utilisation d'informations contenues dans les historiques d'alarmes

Une information importante qui est présente dans un historique d'alarmes et que ne nous n'avons pas pris en compte dans nos travaux est la dimension temporelle. En effet, chaque apparition d'alarme est associée à une date dans l'historique. La prise en compte du temps physique serait utile pour affiner les résultats obtenus. Elle permettrait de traduire les

précédences systématiques au sein d'un réseau de Petri P-temporel. Il serait alors possible de diminuer le langage excédentaire du réseau actuel : les durées minimale et maximale entre les occurrences de deux alarmes liées formeraient les bornes de l'intervalle de temps durant lequel les jetons seraient disponibles pour le tir de la transition en aval de la place qui les contient.

La prise en compte du temps permettrait également d'augmenter la robustesse de la méthode face à des problèmes dus au système d'acquisition tels que la résolution de la gestion des alarmes enregistrées à la même date. Pour résoudre ce problème particulier nous proposons de ne plus considérer les apparitions d'alarmes comme des évènements, mais de s'inspirer de (Saives et al. 2017) en considérant pour chaque date un vecteur évènement où chaque ligne est associée à une alarme. Dans ce cas, pour chaque ligne, la valeur associée à l'évènement serait 0 si l'alarme n'a pas d'occurrence enregistrée à cette date et 1 si elle en a une. Cette méthode permettrait également d'identifier des cas où plusieurs alarmes sont la condition nécessaire à l'occurrence d'une autre.

b. L'utilisation de connaissances sur l'évolution du processus

Parmi les informations dont nous disposons en parallèle de l'historique d'alarmes, certaines peuvent être utilisées pour ajouter des connaissances sur le comportement du système durant l'analyse. Etant donné que les alarmes correspondent à un comportement non désiré du processus industriel, il est naturel de vouloir mettre en parallèle l'historique des fautes et l'historique des alarmes. Cette méthode permettrait de vérifier si des fautes précèdent les alarmes et ainsi de voir si les alarmes sont associées aux mêmes fautes. Cette connaissance peut ainsi être utilisée pour apporter une nouvelle aide à l'analyse de l'expert.

Par exemple, la recherche de précédences systématiques a été utilisée par General Electric Digital pour le compte d'un constructeur automobile. Ce dernier souhaitait améliorer son processus grâce aux données récupérées sur la chaîne de production. Ainsi, par la recherche de précédences systématiques, notre algorithme a permis d'identifier les liens possibles entre une séquence d'actions opérateur, les opérations défectueuses de la ligne de production et les alarmes qu'elles provoquent.

Références bibliographiques

Abele, L, Anic, M, Gutmann, T, Folmer, J, Kleinsteuber, M & Vogel-Heuser, B 2013, 'Combining knowledge modeling and machine learning for alarm root cause analysis.', *IFAC Proceedings Volumes*, vol 46, no. 9, pp. 1843-1848.

Agrawal, R, Imieliński, T & Swami, A 1993, 'Mining association rules between sets of items in large databases', *Acm sigmod record*, vol 22, no. 2, pp. 207-216.

Agrawal, R & Srikant, R 1995, 'Mining sequential patterns', *Eleventh International Conference on Data Engineering (ICDE)*, IEEE, Taipei.

Ahmed, K, Izadi, I, Chen, T, Joe, D & Burton, T 2013, 'Similarity analysis of industrial alarm flood data.', *IEEE Transactions on Automation Science and Engineering*, vol 10, no. 2, pp. 452-457.

Bell, R 1999, 'IEC 61508: Functional safety of electrical/Electronic/programme Electronic safety-related systems: Overview.', IEE Colloquium Controlof Major Accidents and Hazards Directive (COMAH)- Implications for Electrical and Control Enginners.

Cabadi, J-F & Sabot, H 2016, 'Method and system of alarm rationalization in an industrial control system', U.S. Patent Application No. 15/180,134.

Cauvin, S, Cordier, M-O, Dousson, C, Laborie, P, Lévy, F, Montmain, J, Porcheron, M, Servet, I & Travé-Massuyès, L 1998, 'Monitoring and alarm interpretation in industrial environments', *AI Communications*, vol 11, no. 3, pp. 139-173.

Cheng, Y, Izadi, I & Chen, T 2013, 'Pattern matching of alarm flood sequences by a modified Smith–Waterman algorithm.', *Chemical engineering research and design*, vol 91, no. 6, pp. 1085-1094.

Chiang, LH & Braatz, RD 2003, 'Process monitoring using causal map and multivariate statistics: fault detection and identification.', *Chemometrics and intelligent laboratory systems*, vol 65, no. 2, pp. 159-178.

David, R & Alla, H 2005, *Discrete, continuous, and hybrid Petri nets*, Springer, Berlin.

De Souza, JS, Meza, EM, Schilling, MT & Do Coutto Filho, MB 2004, 'Alarm processing in electrical power systems through a neuro-fuzzy approach.', *IEEE Transactions on Power Delivery*, vol 19, no. 2, pp. 537-544.

Dorgo, G & Abonyi, J 2018, 'Sequence mining based alarm suppression.', *IEEE Access*, vol 6, pp. 15365-15379.

Dorgo, G, Varga, K, Haragovics, M, Szabo, T & Abonyi, J 2018, 'Towards Operator 4.0, Increasing Production Efficiency and Reducing Operator Workload by Process Mining of Alarm Data.', *Chemical Engineering Transactions*, vol 70, pp. 829-834.

EEMUA 2007, 'Alarm systems - a guide to design, management and procurement ', Engineering Equipment And Materials Users' Association.

El Sakka, F, Bloch, H, Kinghorst, J, Pirehgalin, MF, Fay, A & Vogel-Heuser, B 2018, 'Improved alarm flood analysis by cluster identification and alarm assignment.', *at-Automatisierungstechnik*, vol 66, no. 8, pp. 647-655.

Folmer, J, Schuricht, F & Vogel-Heuser, B 2014, 'Detection of temporal dependencies in alarm time series of industrial plants.', *IFAC Proceedings Volumes*, vol 47, no. 3, pp. 1802-1807.

Folmer, J & Vogel-Hauser, B 2012, 'Computing dependent industrial alarms for alarm flood reduction.', *International Multi-Conference on Systems, Sygnals & Devices*, IEEE, Chemnitz.

Fullen, M, Schüller, P & Niggemann, O 2017, 'Defining and validating similarity measures for industrial alarm flood analysis.', *15th International Conference on Industrial Informatics*, IEEE, Emden.

Guo, W, Wen, F, Liao, Z, Wei, L & Xin, J 2010, 'An analytic model-based approach for power system alarm processing employing temporal constraint network.', *IEEE Transactions on Power Delivery*, vol 25, no. 4, pp. 2435-2447.

Han, J, Kamber, M & Pei, J 2011, *Data Mining: Concepts and Techniques*, Elsevier.

Han, J, Pei, J & Yin, Y 2000, 'Mining frequent patterns without candidate generation', *ACM sigmod record*, vol 29, no. 2, pp. 1-12.

Health and Safety Executive 1997, 'The explosion and fires at the Texaco Refinery, Milford Haven, 24 July 1994'.

Hopcroft, JE, Motwani, R & Ullman, JD 2003, *Introduction to automata theory, languages, and computation*, 2nd edn, Addison-Wesley, Boston.

International Society of Automation 2009, 'ANSI/ISA-18.2-2016, Management of Alarm Systems for the Process Industries'.

Kinghorst, J, Pirehgalin, MF & Vogel-Heuser, B 2018, 'Graph-based Grouping of Statistical Dependent Alarms in Automated Production Systems.', *IFAC-PapersOnLine*, vol 51, no. 24, pp. 395-400.

Koene, J & Vedam, H 2000, 'Alarm management and rationalization', Third International conference on loss prevention.

Kondaveeti, SR, Izadi, I, Shah, SL, Black, T & Chen, T 2012, 'Graphical tools for routine assessment of industrial alarm systems.', *Computers & Chemical Engineering*, vol 46, pp. 39-47.

Krishnan, NC & Cook, DJ 2014, 'Activity recognition on streaming sensor data', *Pervasive and mobile computing*, vol 10, pp. 138-154.

Laumonier, Y, Faure, J-M, Lesage, J-J & Sabot, H 2017, 'Towards alarm flood reduction"', *22nd International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, Limassol.

Laumonier, Y, Faure, J-M, Lesage, J-J & Sabot, H 2019, 'Discovering Systematic Relations between Alarms for Alarm Flows Reduction', *Conference on Control, Decision and Information Technologies (CoDIT)*, IEEE, Paris.

Liu, X, Noda, M & Nishitani, H 2010, 'Evaluation of plant alarm systems by behavior simulation using a virtual subject.', *Computer & chemical engineering*, vol 34, no. 3, pp. 374-386.

Lucke, M, Chioua, M, Grimholt, C, Hollender, M & Thornhill, NF 2018, 'Online alarm flood classification using alarm coactivations.', *IFAC-PapersOnLine*, vol 51, no. 18, pp. 345-350.

Mackiewicz, RE 2006, 'Overview of IEC 61850 and Benefits.', *IEEE Power Engineering Society General Meeting*, IEEE.

McDonald, JR, Burt, GM & Young, DJ 1992, 'Alarm processing and fault diagnosis using knowledge based systems for transmission and distribution network control.', *IEEE Transactions on Power Systems*, vol 7, no. 3, pp. 1292-1298.

Mikkonen, H, Suojanen, S, Hakkarainen, E & Jegoroff, M 2016, 'Control Method Development Platform for Hybrid Power Plants', *POWER-GEN Europe*, Milan.

Murata, T 1989, 'Petri nets: Properties, analysis and applications', *Proceedings of the IEEE*, vol 77, no. 4, pp. 541-580.

Pétin, J-F, Evrot, D, Morel, G & Lamy, P 2010, 'Combining SysML and formal methods for safety requirements verification.', *22nd International Conference on Software Systems Engineering and their Applications*, Paris.

Saives, J, Faraut, G & Lesage, J-J 2015, 'Identification of discrete event systems unobservable behaviour by petri nets using language projections', *European Control Conference (ECC)*, IEEE, Linz.

Saives, J, Faraut, G & Lesage, J-J 2017, 'Automated partitioning of concurrent discret-event systems for distributed behavioral identification.', *IEEE Transactions on Automation Science and Engineering*, vol 15, no. 2, pp. 832-841.

Schleburg, M, Christiansen, L, Thornhill, NF & Fay, A 2013, 'A combined analysis of plant connectivity and alarm logs to reduce the number of alerts in an automation system.', *Journal of process control*, vol 23, no. 6, pp. 839-851.

Silva, M 1993, 'Introducing Petri nets', *Practice of Petri Nets in manufacturing*, pp. 1-62.

Simeu-Abazi, Z, Lefebvre, A & Derain, J-P 2011, 'A methodology of alarm filtering using dynamic fault tree.', *Reliability Engineering & System Safety*, vol 96, no. 2, pp. 257-266.

Thornhill, NF, Cox, JW & Paulonis, MA 2003, 'Diagnosis of plant-wide oscillation through data-driven analysis and process understanding.', *Control Engineering Practice*, vol 11, no. 12, pp. 1481-1490.

Vasquez Capacho, JW, Subias, A, Travé-Massuyès, L & Jimenez, F 2017, 'Alarm management via temporal pattern learning.', *Engineering Applications of Artificial Intelligence*, vol 65, pp. 506-516.

Wang, J & Chen, T 2014, 'An online method to remove chattering and repeating alarms based on alarm durations and intervals', *Computers & Chemical Engineering*, vol 67, pp. 43-52.

Wang, J, Li, H, Huang, J & Su, C 2015, 'A data similarity based analysis to consequential alarms of industrial processes.', *Journal of Loss Prevention in the Process Industries*, vol 35, pp. 29-34.

Wang, J, Xu, J & Zhu, D 2014, 'Online root-cause analysis of alarms in discrete Bayesian networks with known structures.', *Proceeding of the 11th World Congress on Intelligent Control and Automation*, IEEE, Shenyang.

Wang, J, Yang, F, Chen, T & Shah, SL 2016, 'An overview of industrial alarm systems: Main causes for alarm overloading, research status, and open problems.', *IEEE Transactions on Automation Science and Engineering*, vol 13, no. 2, pp. 1045-1061.

Wei, L, Guo, W, Wen, F, Ledwich, G, Liao, Z & Xin, J 2011, 'An online intelligent alarm-processing system for digital substations.', *IEEE transactions on Power Delivery*, vol 26, no. 3, pp. 1615-1624.

Wen, FS & Chang, CS 1997, 'Tabu search approach to alarm processing in power systems.', *IEE Proceedings-Generation, Transmission and Distribution*, vol 144, no. 1, pp. 31-38.

Wuhib, F, Dam, M, Stadler, R & Clemm, A 2005, 'Decentralized Computation of Threshold Crossing Alerts', *16th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management*, Springer, Barcelona, Spain.

Wunderlich, P & Niggemann, O 2017, 'Structure learning methods for Bayesian networks to reduce alarm floods by identifying the root cause.', *22nd International Conference on Emerging Technologies and Factory Automation*, IEEE, Limassol.

Xu, S, Adhitya, A & Srinivasan, R 2014, 'Hybrid mode-based framework for alarm anticipation', *Industrial & Engineering Chemistry Research*, vol 53, no. 13, pp. 5182-5193.

Yang, F, Shah, SL & Xiao, D 2010, 'Correlation analysis of alarm data and alarm limit design for industrial processes', *Proceedings of the 2010 American Control Conference*, IEEE, Baltimore.

Yang, F, Shah, SL, Xiao, D & Chen, T 2012, 'Improved correlation analysis and visualization of industrial alarm data.', *ISA transactions*, vol 51, no. 4, pp. 499-506.

Yang, Z, Wang, J & Chen, T 2013, 'Detection of correlated alarms based on similarity coefficients of binary data.', *IEEE Transactions on Automation science and Engineering*, vol 10, no. 4, pp. 1014-1025.

Zhao, Q & Bhowmick, SS 2003, 'Sequential pattern mining: A survey', ITechnical Report CAIS Nayang Technological University Singapore.

université PARIS-SACLAY | ÉCOLE DOCTORALE Sciences et technologies de l'information et de la communication (STIC)

Titre : Apport des paradigmes des Systèmes à Evènements Discrets pour la réduction du flux d'alarmes industrielles

Mots clés : alarmes industrielles, Systèmes à Evènements Discrets, filtrage d'alarmes, recherche de motifs, réseaux de Petri

Résumé : Les systèmes d'alarmes jouent un rôle critique dans la bonne exploitation des installations industrielles modernes. Cependant, dans la plupart de ces systèmes, les alarmes ne sont pas toujours traitées correctement par les opérateurs car il y a régulièrement beaucoup trop d'alarmes à gérer, notamment lors des avalanches d'alarmes.

Pour réduire le flux d'alarmes, notre approche consiste à détecter des redondances entre alarmes qui pourraient être supprimées. Pour atteindre cet objectif, nous recherchons dans un premier temps les adjacences fréquentes entre les alarmes contenues dans un historique. Ceci est réalisé en adaptant l'algorithme de recherche de motifs fréquents *AprioriAll*. Nous explorons également une seconde méthode consistant à trouver des précédences systématiques. Pour les découvrir, nous identifions les relations de domination et de mutuelle dépendance contenues dans l'historique des alarmes. Pour faciliter l'analyse experte, les relations découvertes sont traduites sous la forme d'un réseau de Pétri.

Ces deux méthodes sont ensuite confrontées à un historique d'alarmes industriel fourni par General Electric. Les résultats obtenus montrent que nos deux méthodes permettent une réduction globale du flux d'alarmes qui est plus importante durant les avalanches.

Title: Contribution of Discrete Event Systems paradigms for reducing industrial alarm flows

Keywords: Industrial alarms, Discrete Event Systems, alarm filtering, pattern mining, Petri nets

Abstract: Alarm systems play an important role for the safe and efficient operation of modern industrial plants. However, in most of industrial alarm systems, alarm flows cannot always be correctly managed by the operators as they often turn into alarm floods, sequences of numerous alarms occurring in a short period of time.

To reduce the alarm flows, this report focuses on detection of redundant alarms that could be removed. This objective is met by, first, looking for frequent adjacency in the alarm log. To identify them, the frequent pattern mining algorithm *AprioriAll* is adapted. Another way to find potentially redundant alarms is to look for systematic predecessors. To discover them, dominations and mutual dependencies contained in the alarm log are identified. To ease this analysis, the discovered relations are depicted in the form of Petri nets.

Both those methods are then tested against an industrial alarm log made available by General Electric. The results show that both methods allow a reduction of the alarm flow, with the biggest reduction being during alarm floods.