Figure: Swissgrid

Cyberattaques sur les réseaux électriques

Détection d'anomalies IT par machine learning | Lors de cyberattaques de tromperie, le réseau IT remonte des informations erronées sur l'état opérationnel d'un réseau électrique. Ceci peut provoquer des réactions inadéquates de la part de l'opérateur, pouvant mener, dans un cas extrême, à un blackout. De telles attaques peuvent toutefois être détectées par des algorithmes de machine learning.

GUILLAUME DUBUIS, PHILIPPE JACQUOD

Les réseaux électriques sont des systèmes cyber-physiques. Ils sont formés d'un réseau physique – comprenant notamment les centrales de production ainsi que les lignes et transformateurs qui les relient aux consommateurs – et d'un réseau IT. Ce dernier transmet, d'une part, les informations sur l'état opérationnel du réseau physique – niveaux de tension, fréquence AC, puissances actives et réactives injectées ou soutirées, etc. – à l'opérateur et, d'autre part, les commandes de ce dernier vers les divers composants du réseau physique.

Ces systèmes sont actuellement soumis à rude épreuve. Tout d'abord, parce que la sortie annoncée du nucléaire, l'électrification croissante des transports et de la production de chaleur, ainsi que la montée en puissance des nouvelles sources d'énergie renouvelables les poussent à fonctionner dans des modes opérationnels nouveaux, souvent plus proches de leurs limites de capacité. Ensuite, parce que leur numérisation toujours croissante – leur transformation en « smart grids » – multiplie les composants électroniques nécessaires à leur fonctionnement. Ces composants représentant autant de portes d'entrée vers le réseau IT, d'évidents problèmes de cybersécurité se posent. De fait, la pression induite par des cyberattaques sur les infrastructures d'importance stratégique ne cesse d'augmenter et rend ces dernières vulnérables [1].

Cette « épée de Damoclès » n'est pas acceptable à long, ni même à moyen terme, sur une infrastructure critique d'importance stratégique. Des solutions efficaces doivent être trouvées aux problèmes de cybersécurité qui pèsent sur les réseaux électriques. Le présent article résume les résultats obtenus dans le cadre d'un projet, financé par armasuisse et réalisé à la HES-SO Valais, portant sur la détection de cyberattaques sur le réseau IT de systèmes électriques. Celui-ci a pour objectif de développer et de tester des algorithmes de détection d'anomalies dans les données d'état opérationnel du réseau reçues par l'opérateur.

Pourquoi utiliser des algorithmes de machine learning?

Considérons un cyberattaquant ayant pénétré le système Scada d'une centrale de production. Ceci lui permet d'en-

voyer des informations erronées sur le statut de la centrale. Par exemple, cette dernière est annoncée à l'arrêt, alors qu'en réalité, elle produit à sa puissance nominale – ou vice-versa. En soi, les deux états sont parfaitement acceptables – il n'est pas anormal qu'une centrale produise, ni qu'elle soit à l'arrêt. Il s'agit donc de détecter des anomalies qui n'en sont que d'un point de vue contextuel, c'est-à-dire du point de vue de l'état opérationnel global du système.

De telles anomalies peuvent en principe être détectées par des méthodes traditionnelles de génie électrique. L'approche standard basée sur les flux de puissance (power flow) indique un manque de production non annoncé, ou une production existante mais annoncée manquante: le calcul ne converge pas. La procédure doit néanmoins être complétée afin de pouvoir localiser les centrale(s), station(s) ou ligne(s) dont les données transmises à l'opérateur sont erronées. Ceci peut se révéler coûteux en temps de calcul et difficilement implémentable en temps réel, alors même que les nouveaux modes de fonctionnement des réseaux requièrent de plus en plus de flexibilité et des temps de réaction plus courts de la part de tous les acteurs. Il est donc impératif de pouvoir analyser les données reçues par l'opérateur sur l'état du réseau plus fréquemment, plus rapidement et de manière fiable.

Dans le cadre de ce projet, il a été décidé d'explorer une nouvelle direction en suivant une approche basée sur des algorithmes de machine learning. L'un des résultats principaux du projet – la démonstration que de tels algorithmes, fonctionnant purement sur des données, peuvent détecter de manière fiable des anomalies rares et contextuelles pour autant qu'ils soient bien entraînés – est brièvement discuté ci-après. De par leur rapidité d'exécution, ces algorithmes semblent prometteurs pour l'analyse des données opérationnelles de réseaux électriques ainsi que pour la détection éventuelle d'anomalies dans ces données.

Génération d'un modèle synthétique du réseau

Dans ce projet, des réseaux de transport à très haute tension, dont les données opérationnelles réelles sont notoirement difficiles à obtenir, ont été considérés. Il a donc fallu développer un modèle synthétique du réseau suisse à très haute tension opéré par Swissgrid, en incluant quelques lignes et stations se trouvant à la périphérie du pays. La topologie du réseau est illustrée dans la **figure 1.**

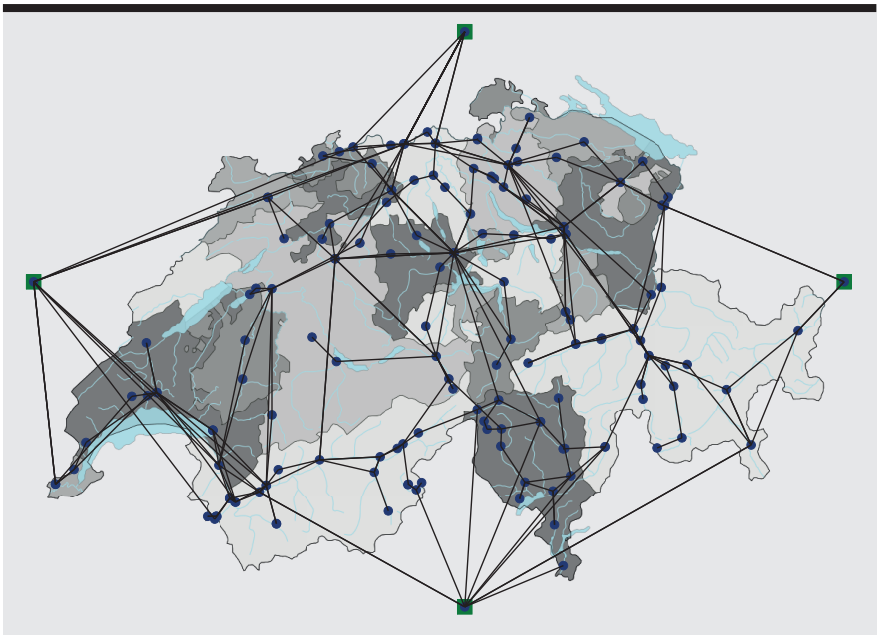

Figure 1 Topologie du réseau électrique à très haute tension étudié (modèle synthétique).

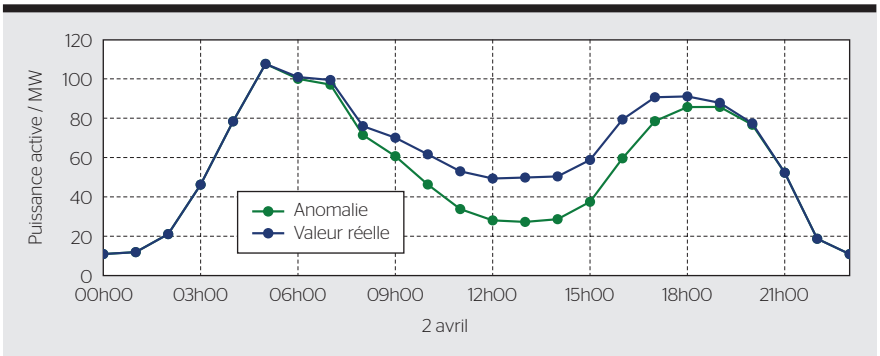

Figure 2 Anomalie de distorsion, générée pour une centrale hydroélectrique de retenue.

La phase d'entraînement des algorithmes de machine learning est gourmande en données de vérité terrain (ground truth). Il est donc indispensable de disposer d'un ensemble de données de vérité terrain aussi grand que possible. Dans le cas considéré, il s'agit de séries temporelles de productions et de soutirages de puissance électrique sur chaque bus du réseau. Tout comme pour le modèle de réseau lui-même, ces données ont été générées synthétiquement, à partir d'un flux de puissance optimal (optimal power flow, OPF) au niveau européen basé sur les données de consommation nationales. Les détails du modèle sont discutés dans les articles [2] et [3]. Ce modèle synthétique fournit une vérité terrain sous la forme d'un jeu de données d'une année avec résolution horaire pour toutes les productions et consommations suisses, ainsi que le flux entre la Suisse et les pays voisins. La majorité de ce jeu de données est utilisée pour l'entraînement des algorithmes. L'ensemble test contient 20 % du jeu de données complet.

Entraînement et tests des algorithmes

Tant pour l'entraînement que pour la phase de test, des anomalies doivent être insérées dans la vérité terrain. La forme de ces anomalies est en principe arbitraire, et trois types d'anomalies ont été considérés, tous suffisamment réalistes pour ne pas directement attirer l'attention de l'opérateur. Premièrement, de nombreuses centrales fonctionnent en mode « on/off », pro-

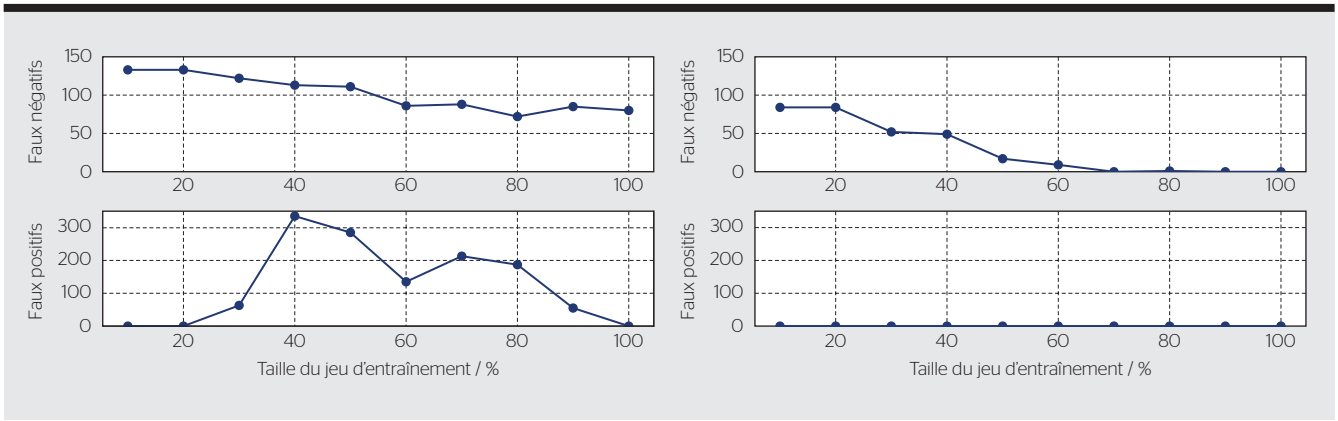

Figure 3 Performance de l'algorithme MLPC pour une cyberattaque de distorsion sur la production d'une centrale hydroélectrique de retenue (à gauche) et au fil de l'eau (à droite).

duisant à leur puissance nominale, ou pas du tout. Pour ces cas, les anomalies annoncent la centrale en production, alors qu'elle ne l'est pas, ou vice-versa. Deuxièmement, des anomalies de distorsion ont été considérées, où les données de production sont modifiées de manière lisse, comme illustré dans la **figure 2** pour une centrale hydroélectrique de retenue. Troisièmement, des anomalies de rediffusion (data replay) ont été générées, où un attaquant a enregistré une série temporelle de production pour la rejouer plus tard. Les anomalies « on/off » étant relativement faciles à détecter, la suite de cet article se focalise sur les anomalies de distorsion et de rediffusion.

Afin de pouvoir évaluer la performance des algorithmes, quatre centrales hydroélectriques ont été sélectionnées – deux centrales au fil de l'eau et deux centrales de retenue –, dont la production est modifiée environ un jour sur dix de manière lisse. L'objectif consiste à détecter ces anomalies aussi vite que possible. Il est évident que la détection ne peut se faire instantanément – des anomalies lisses mettent un certain temps à se développer et donc à différer significativement de la courbe de puissance réelle **(figure 2)**. La performance d'un algorithme est mesurée par le nombre d'anomalies non détectées (faux négatifs), le nombre de fausses alertes (faux positifs) et le temps moyen mis pour la détection.

Enfin, quatre différents algorithmes ont été entraînés et testés – *k* plus proches voisins, machines à vecteurs de support, forêt d'arbres décisionnels et réseau de neurones multicouches (MLPC). C'est ce dernier algorithme qui a donné les meilleurs résultats, et ceux-ci sont expliqués plus en détail ci-après.

Jours détectés	Centrale de retenue n° 66	Centrale de retenue n° 127	Centrale au fil de l'eau n° 43	Centrale au fil de l'eau n° 57
7	-	1,6 ts - 16,1 % - 4,7 MW	0,7 ts - 7,4 % - 4,7 MW	0,9 ts - 7,6 % - 1,8 MW
5	3,4 ts - 23,3 % - 37,0 MW	2,7 ts - 16,4 % - 4,4 MW	0,7 ts - 7,4 % - 4,7 MW	0,9 ts - 7,6 % - 1,8 MW

Tableau 1 Performances de l'algorithme MLPC en fonction du nombre de détections pour des anomalies de distorsion de 50 % d'amplitude relative sur les quatre centrales de production étudiées. Le jeu de données test est constitué de 7 jours attaqués.

Jours détectés	Centrale de retenue n° 66	Centrale de retenue n° 127	Centrale au fil de l'eau n° 43	Centrale au fil de l'eau n° 57
7	3,3 ts - 30,0 % - 14,8 MW	3,2 ts - 49,1 % - 14,1 MW	N/A	N/A
6	2,6 ts - 32,5 % - 17,5 MW	3,2 ts - 49,1 % - 14,1 MW	0,3 ts - 3,7 % - 2,5 MW	N/A
5	3,7 ts - 162,5 % - 34,0 MW	2,7 ts - 45,1 % - 16,7 MW	0,3 ts - 3,7 % - 2,5 MW	0,6 ts - 4,6 % - 1,3 MW

Tableau 2 Performances de l'algorithme MLPC en fonction du nombre de détections pour des anomalies de rediffusion sur les quatre centrales de production étudiées. Le jeu de données test est constitué de 7 jours avec anomalies pour les centrales n° 66 et 127, de 6 jours pour la centrale n° 43 et de 5 jours pour la n° 57.

Influence de la taille de l'ensemble d'entraînement

Dans un premier temps, les travaux se sont concentrés sur l'évolution des performances de l'algorithme MLPC en fonction de la taille de l'ensemble d'entraînement. Les sept anomalies considérées couvrent un total de 100 à 130 pas de temps, et la **figure 3** montre le nombre de faux positifs et de faux négatifs. La détection de ces anomalies sur une centrale au fil de l'eau est parfaite avec un entraînement sur 90 % du jeu de données de vérité terrain. Ceci n'est pas inattendu, la production de ces centrales étant presque constante sur une journée. La détection est plus difficile sur une centrale de retenue, à la production beaucoup plus fluctuante. Néanmoins, les résultats présentés dans la **figure 3** suggèrent qu'un entraînement suffisamment long permettrait d'atteindre des performances satisfaisantes.

Reconnaissance des anomalies à temps

Pour qu'un algorithme soit considéré comme efficace, il n'est néanmoins pas nécessaire que tous les pas de temps d'une anomalie soient identifiés. Les anomalies lisses considérées mettent du temps à se développer et ne sont pas problématiques du point de vue de la sécurité du réseau tant que la puissance annoncée reste proche de la production réelle. La détection est donc considérée comme couronnée de succès si une anomalie est détectée suffisamment tôt, c'est-à-dire avant que la puissance annoncée diffère significativement de la puissance réelle.

Les mesures de performance des algorithmes développés au cours de ce projet sont présentées pour des cyber-

Figure: HES-SO Valais

attaques de distorsion et de rediffusion dans les **tableaux 1 et 2** respectivement. Les performances sont mesurées en fonction du nombre d'anomalies détectées, du temps moyen nécessaire pour les détecter (en nombre de pas de temps, indiqué par « ts » pour « time step ») ainsi que des déviations absolue et relative (en pourcent) entre les puissances réelle et annoncée au moment de la détection. Des quatre centrales de production considérées, deux sont des centrales de retenue (n° 66 et 127), et deux des centrales au fil de l'eau (n° 43 et 57).

On remarque premièrement que seules quelques attaques ne sont pas détectées comme telles, et ce, uniquement pour la centrale de retenue n° 66, particulièrement fluctuante. Deuxièmement, les anomalies sont détectées rapidement, après quelques pas de temps seulement. La déviation relative de la production annoncée atteint au maximum 7,6 % pour les centrales au fil de l'eau. Pour les centrales à retenue, les déviations peuvent se superposer à des productions très faibles, ce qui augmente la déviation relative. Dans tous les cas, des déviations correspondant à moins de 1 % de la puissance totale produite/consommée sur le réseau suisse sont déjà détectées, donc bien avant que l'attaque ne puisse être véritablement problématique.

Perspectives

Ce bref compte rendu montre le potentiel du machine learning dans l'aide à la détection de cyberattaques de tromperie annonçant des données opérationnelles fausses à un opérateur de réseau électrique. Seuls les résultats de l'algorithme le plus prometteur – un MLPC – ont été présentés. Tout indique qu'un renforcement de la phase d'entraînement permettrait d'atteindre des résultats d'alerte fiables – identifiant toutes les attaques avec relativement peu de fausses alertes. Ces résultats ne sont pas spécifiques au modèle considéré, d'autres résultats de qualité comparable ayant été obtenus pour d'autres modèles de réseau, en particulier ceux de SimBench [4].

Le projet présenté dans cet article est toujours en cours de finalisation, et il devrait être possible de parfaire les performances de l'algorithme MLPC utilisé en augmentant la taille du jeu de données d'entraînement. La difficulté consiste à générer des données synthétiques réalistes, en particulier sans corrélations excessives pouvant favoriser artificiellement les algorithmes de machine learning. L'accès à des données réelles permettrait de contourner cette difficulté, le but ultime étant de déployer les algorithmes de détection développés en situation et en temps réels.

Finalement, l'angle proposé dans cet article a été celui de la cybersécurité. À noter toutefois que cet algorithme détecte des anomalies, quelle que soit leur origine. En particulier, des erreurs techniques de transmission de données sont tout aussi bien détectées que celles résultant d'une cyberattaque.

Références

[1] M. Galus, F. Heymann, S. Henry, « Cyber-Sicherheit und Cyber Resilienz für die Schweizer Stromversorgung », OFEN, 2021.

[2] M. Tyloo, L. Pagnier, P. Jacquod, « The Key Player Problem in Complex Oscillator Networks and Electric Power Grids : Resistance Centralities Identify Local Vulnerabilities », Science Advances, Vol. 5, p. eaaw8359, 2019. doi.org/10.1126/sciadv.aaw8359

[3] L. Pagnier, P. Jacquod, « Inertia Location and Slow Network Modes Determine Disturbance Propagation in Large Scale Power Grids », Plos One, Vol. 14, p. e0213550, 2019. doi.org/10.1371/journal.pone.0213550

[4] S. Meinecke, D. Sarajlić, S. Drauz, A. Klettke, L.-P. Lauven, C. Rehtanz, A. Moser, M. Braun, « SimBench – A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis », Energies, 2020. doi.org/10.3390/en13123290

Auteurs

Guillaume Dubuis est étudiant en Master à la HES-SO Valais et Master Thesis Fellow au Cyber-Defence Campus Armasuisse/EPFL.
→ HES-SO Valais, 1950 Sion
→ guillaume.dubuis@hevs.ch

Philippe Jacquod est professeur à la HES-SO Valais, où il dirige une équipe de recherche active dans le domaine des réseaux électriques.
→ philippe.jacquod@hevs.ch

IN KÜRZE

Cyberangriffe auf Stromnetze

Erkennung von IT-Anomalien durch Machine Learning

Bei Täuschungsangriffen im Internet gibt das IT-Netzwerk falsche Informationen über den Betriebszustand eines Stromnetzes weiter. Dies kann zu Fehlreaktionen des Betreibers führen, die im Extremfall einen Blackout verursachen können. Solche Angriffe können jedoch mit Machine-Learning-Algorithmen erkannt werden, wie ein Projekt an der HES-SO Wallis gezeigt hat.

Im Projekt wurden Algorithmen zur Erkennung von Anomalien in den vom Netzbetreiber erhaltenen Daten über den Betriebszustand des Netzes entwickelt und getestet. Da echte Betriebsdaten schwer zu beschaffen sind, wurde ein Modell des Schweizer Swissgrid-Höchstspannungsnetzes synthetisch erstellt und Zeitreihen der Stromerzeugung und -entnahme auf jedem Bus des Netzes generiert.

In diese Datensätze wurden verschiedene Arten von Anomalien eingefügt: unter anderem Verzerrungsanomalien, bei denen die Produktionsdaten verändert wurden, und Replay-Anomalien, bei denen ein Angreifer eine Produktionszeitreihe aufzeichnete, um sie später wieder abzuspielen. Dann wurden vier verschiedene Algorithmen trainiert und mit den Produktionsdaten von zwei Laufwasser- und zwei Speicherkraftwerken getestet.

Der MLPC-Algorithmus (Multilayer Neuronal Network) lieferte die besten Ergebnisse, insbesondere mit einer perfekten Erkennung von Verzerrungsanomalien bei der Produktion eines Laufwasserkraftwerks, sofern die Grösse des Datensatzes für das Training ausreichend gross ist. Nur wenige Angriffe wurden nicht als solche erkannt, und das auch nur bei einem besonders schwankenden Speicherkraftwerk. In allen Fällen werden bereits Abweichungen festgestellt, die weniger als 1 % der gesamten im Schweizer Netz erzeugten/verbrauchten Leistung entsprechen, also lange bevor der Angriff wirklich problematisch sein kann. Ausserdem deutet alles darauf hin, dass eine verstärkte Trainingsphase zu zuverlässigen Warnergebnissen führen würde, die alle Angriffe mit relativ wenigen Fehlalarmen identifizieren würden.