Cybersécurité en réalité virtuelle : améliorer le processus de détection d'intrusion, d'investigation et de décision via l'utilisation de techniques de visualisations 3D immersives

Nicolas Delcombel

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE
MINES-TÉLÉCOM ATLANTIQUE BRETAGNE
PAYS DE LA LOIRE – IMT ATLANTIQUE

ÉCOLE DOCTORALE N° 648
Sciences pour l'Ingénieur et le Numérique
Spécialité : *Informatique*

Par
Nicolas DELCOMBEL

Cybersécurité en Réalité Virtuelle

Améliorer le processus de détection d'intrusion, d'investigation et de décision via l'utilisation de techniques de visualisations 3D immersives

Thèse présentée et soutenue à IMT Atlantique, Rennes, le 19 décembre 2023
Unité de recherche : LabSTICC (UMR CNRS 6285)
Thèse N° : 2023IMTA0387

Rapporteurs avant soutenance :

Pierre PARREND — Professeur
EPITA Strasbourg
Marcos SERRANO — Maître de Conférences
Université Toulouse 3

Composition du Jury :

Présidente :	Françoise SAILHAN	Professeure IMT Atlantique campus de Brest
Examinateurs :	Pierre PARREND	Professeur EPITA Strasbourg
	Marcos SERRANO	Maître de Conférences Université Toulouse 3
	Arnaud PROUZEAU	ISFP (Inria Starting Faculty Position) INRIA Bordeaux Sud-Ouest
Dir. de thèse :	Thierry DUVAL	Professeur IMT Atlantique campus de Brest
Encadrant de thèse :	Marc-Oliver PAHL	Directeur d'étude IMT Atlantique campus de Rennes

Remerciements

Je souhaite exprimer ma gratitude envers toutes les personnes qui ont contribué à la réalisation de ces travaux, en premier lieu mes codirecteurs de thèse Thierry Duval et Marc Oliver Pahl, pour leur implication, leurs conseils judicieux et leur bienveillance. Mes remerciements vont également à Marcos Seranno et Pierre Parrend qui ont généreusement accepté de relire ce manuscrit et m'ont prodigué des conseils précieux pour son amélioration. Je remercie également François Sailland et Arnaud Prouzeau pour leurs retours sur mes travaux. Je tiens à adresser mes remerciements à l'ensemble du département informatique de l'IMT Atlantique, campus de Brest, qui a permis de rendre cette période de thèse plus agréable. En particulier, je remercie mes collègues du bureau D03-115A pour leur bonne humeur ainsi que leur bon humour. Enfin, tout ce travail n'aurait été possible sans les encouragements de mes ami.e.s que je remercie chaleureusement. Et évidement, je tiens à terminer par un grand merci à mes parents et à ma sœur pour leur soutien indéfectible.

Table des matières

TABLE DES FIGURES

Liste des tableaux

Abréviations

IDS : **I**ntrusion **D**etection **S**ystem ; système de détection d'intrusion

NIDS : **N**etwork **I**ntrusion **D**etection

Software ; système de détection d'intrusion réseaux

SIEM : **S**ecurity **I**nformation and **e**vent **M**anagement ; système de gestion des informations et des événements de sécurité

RV : **R**éalité **V**irtuelle

WIMP : **W**indows, **I**cons, **M**enus and **P**ointing device ; fenêtres, icônes, menus et dispositif de pointage

INTRODUCTION

Contexte : les signaux faibles en cybersécurité

La cybersécurité représente un élément incontournable pour la technologie moderne, jouant un rôle indispensable dans la préservation de la sécurité des informations et des systèmes sensibles. C'est une préoccupation majeure pour toutes les organisations à l'échelle mondiale. Face à la recrudescence constante du nombre d'attaques au cours des dernières années, ces grandes organisations s'organisent afin de pouvoir contrer efficacement ces menaces. Pour cela, elles mettent en place des équipes spécialisées en cybersécurité, notamment des équipes dédiées à la défense en temps réel de leurs actifs contre une variété de menaces. Ces équipes rassemblent divers experts en cybersécurité, habituellement regroupés au sein de Centres des Opérations de Sécurité (SOC, pour Security Operation Center).

Ces analystes de la cybersécurité qui travaillent au sein des SOCs doivent traiter des alertes levées par des logiciels de détection d'intrusion (IDS). Ceux-ci lèvent souvent un grand nombre de faux positifs et les analystes peuvent avoir du mal à traiter correctement les alertes au vu du grand volume de données auquel ils font face. Les opérateurs utilisent différents indices pour répondre à ces alertes, comme l'identité d'un utilisateur ayant accédé à des données sensibles ou un nombre trop important de connexion vers un serveur extérieur. Seuls, ces indices ne sont pas des signaux alarmants, mais la présence de plusieurs d'entre eux peut être le signe d'une attaque. Ce type d'indice correspond à la définition donnée des signaux faibles dans le domaine de la criminalité : "des avertissements diffus et difficiles à détecter qui nécessitent la corrélation de plusieurs indicateurs pour indiquer une activité [cyber] criminelle" [1]. Ainsi, le but des opérateurs est de corréler ces signaux faibles pour répondre aux alertes levées automatiquement.

Pour réaliser leurs tâches, les opérateurs des SOCs utilisent habituellement des interfaces WIMP (Windows, Icons, Menus and Pointing device pour « fenêtres, icônes, menus et dispositif de pointage) ainsi que des lignes de commandes et des visualisations 2D. Ces dernières sont limitées en termes de place d'affichage ainsi qu'en moyens d'interactions. Des représentations 3D immersives des données pourraient aider à surveiller plus

efficacement ces systèmes. En effet, le domaine de l'Immersive Analytics a été défini en 2015 [2] et cherche à déterminer (entre autres) comment utiliser les technologies immersives pour améliorer les visualisations de données afin d'aider à de meilleures prises de décisions. Même si l'Immersive Analytics n'est pas encore utilisé dans le domaine de la cybersécurité, il a déjà prouvé son utilité dans d'autres domaines similaires comme la médecine ou l'architecture. Ainsi, mes travaux de thèse portent sur l'usage de technologies de visualisation immersive afin de faciliter la détection et la gestion des signaux faibles en cybersécurité. La première question de recherche de ces travaux est la suivante : **comment l'utilisation de l'Immersive Analytics peut-elle permettre aux opérateurs des SOCs d'être plus efficients dans le traitement des signaux faibles ?** Par la suite, l'étude de l'état de l'art des visualisations va en effet nous montrer que les visualisations 2D actuelles ont des limitations que les visualisations de données 3D peuvent dépasser, ce qui nous conduira à proposer un prototype de visualisation immersive plus efficace que les visualisations 2D de l'état de l'art. Durant nos expériences avec des experts du domaine, nous remarquerons que même si les analystes de la cybersécurité entrevoient le potentiel des représentations 3D, ils ne les utilisent pas actuellement dans leur travail. Cela nous conduira à notre deuxième question de recherche : **quels sont les verrous actuels à l'adoption de l'Immersive Analytics par les opérateurs cyber, et comment les débloquer ?**

Ces travaux s'inscrivent dans le cadre de la Chaire Cyber-CNI de l'Institut Mines-Télécom, portée par IMT Atlantique. Cette chaire fait de la recherche et de la formation dans le domaine de la cybersécurité des infrastructures critiques (réseaux d'énergie, processus industriels, usines de production d'eau, systèmes financiers, etc.). Elle collabore étroitement avec Télécom ParisTech, Télécom SudParis et les entreprises Airbus Defence and Space, AMOSSYS, BNP Paribas, EDF et Nokia Bell Labs. La chaire est financée par le Conseil Régional de Bretagne, par des fonds européens FEDER et par ses partenaires industriels.

Structure et contributions

Dans ce manuscrit, les similitudes entre différentes visualisations de cybersécurité de la littérature seront décrites en utilisant la typologie de Munzner [3] qui permet de décrire les systèmes de visualisation en fonction des données visualisées, des tâches remplies par l'utilisateur, et des techniques de visualisations misent en place. Cette analyse révélera

les exigences communes pour pouvoir réaliser de bonnes visualisations de cybersécurité et identifiera les cas d'utilisation où l'Immersive Analytics est plus performante que les analyses à base de visualisations en 2D. Les deux principaux avantages de l'utilisation de l'Immersive Analytics pour la visualisation des données de cybersécurité que nous mettrons en avant sont de pouvoir afficher le temps et la corrélation entre vues en utilisant trois dimensions. Basé sur ces avantages, nous présenterons un prototype 3D qui sera testé par des utilisateurs novices et experts sur un scénario de réponse à alertes basé sur un jeu de données de la littérature, et le retour d'information soulignera l'importance de l'accès multivues et de l'intégration des outils classiques de la cybersécurité dans l'environnement de visualisation. Nous présenterons ensuite une preuve de concept qui intègre les outils de la cybersécurité dans un environnement immersif pour supporter le workflow des opérateurs cyber.

Ce manuscrit est organisé de la façon suivante : dans le premier chapitre, nous mènerons une analyse du travail effectué dans les SOCs. Les opérateurs cyber qui y travaillent utilisent des systèmes de détection d'intrusion (IDS) qui permettent une approche automatique et rapide pour détecter des intrusions et y répondre. Ces IDS soulèvent néanmoins de nombreux faux positifs. Ainsi, le travail des opérateurs est bien davantage de trier les alertes levées par ces IDS que de surveiller le réseau. Pour répondre à ces alertes, ils ont besoin de comprendre leur contexte le plus rapidement possible. Afin d'accélérer l'acquisition du contexte des alertes, la recherche en cybersécurité s'est tournée vers la visualisation pour transmettre aux opérateurs l'état du système le plus rapidement possible.

Dans le deuxième chapitre, nous montrerons que les visualisations en cybersécurité ont de nombreuses caractéristiques communes. Pour cela, nous avons effectué une analyse complète de la littérature existante à l'aide du cadre théorique proposé par Munzner, qui a l'intérêt de prendre en compte le **but** de l'utilisateur de la visualisation. Cette analyse permet d'identifier les points communs des visualisations étudiées, ainsi que les exigences auxquelles doivent répondre les visualisations de cybersécurité efficaces. Nous identifierons principalement l'utilisation de la troisième dimension pour représenter le temps, particulièrement au sein des réseaux, ainsi que pour la création d'un "Espace à penser", c'est-à-dire utiliser l'espace de l'environnement de visualisation pour visualiser et organiser les différentes étapes des raisonnements des opérateurs.

Dans le troisième chapitre, nous utiliserons cette catégorisation du corpus de la visualisation pour la cybersécurité pour identifier les cas d'usage proches dans l'état de l'art en Immersive Analytics. Ainsi, nous identifierons les cas d'usage où des visualisations

basées sur l'Immersive Analytics s'avèrent plus efficaces que les visualisations 2D. Cela nous permettra de détecter les points sur lesquels l'Immersive Analytics manque encore de solutions pour répondre aux exigences de la cybersécurité. Nous avons ainsi identifié deux avantages clés de l'utilisation de l'Immersive Analytics pour la visualisation des données de cybersécurité, grâce à l'utilisation de la troisième dimension qui permet d'une part d'afficher le temps et d'autre part de corréler les données plus efficacement entre elles qu'en 2D.

Dans le quatrième chapitre, pour démontrer ces avantages, nous avons développé un prototype 3D hélicoïdal de visualisation de données temporelles qui permet de corréler les motifs périodiques au sein des données avec des alertes levées par un algorithme de détection. Nous présenterons les expériences utilisateurs menés sur ce prototype avec des utilisateurs novices et par des utilisateurs experts lors d'une expérimentation. Les résultats montrent que lorsque les données deviennent plus complexes, la solution 3D est significativement plus performante que la solution 2D. De plus, la visualisation 3D permet de générer davantage d'engagement dans la tâche de la part de l'utilisateur au prix d'un léger effort physique supplémentaire.

Enfin, dans un cinquième chapitre, nous nous intéresserons aux retours utilisateurs de l'expérience précédente. Ceux-ci soulignent l'importance de l'accès à plusieurs vues synchronisées et aux outils métiers dans l'environnement de visualisation, ce qui concorde avec les exigences trouvées dans le troisième chapitre. En conséquence, nous présenterons un concept d'environnement immersif qui permet d'intégrer des outils d'agrégation et d'analyse de données souvent utilisés en cybersécurité avec des visualisations 3D. Cela permet de faire le lien entre le recueil de données et les outils métiers, leurs analyses, et les visualisations sans interruption dans le travail des opérateurs.

Ainsi, les contributions de cette thèse sont :

— une étude de la littérature qui permet de décrire les systèmes de visualisations à destination des SOCS leur flux opérationnel. Cette étude met en valeur les manquements des systèmes actuels pour permettre à la communauté de recherche en visualisation de s'approprier les spécificités des challenges de la cybersécurité ;
— la proposition d'un agenda de recherche pour l'usage de l'Immersive Analytics dans les SOCs pour permettre aux chercheurs en Immersive Analytics d'utiliser les use-cases de la cybersécurité et pour permettre aux chercheurs en visualisation pour la cybersécurité d'avoir accès à des visualisations 3D efficaces ;
— Un concept de visualisation immersive permettant de corréler les données issues

de la cybersécurité. Ce concept a été instancié et évalué sur un scénario de cybersécurité afin de faire des recommandations pour la conception de visualisations immersives, en particulier dans le domaine de la cybersécurité ;

— Un concept de workspace immersif pouvant afficher des visualisations 3D afin de montrer qu'il est possible de supporter un workflow de cybersécurité au sein d'une solution immersive.

Chapitre 1

Contexte : Cybersécurité des SOCs

La sécurité des systèmes d'information vise à protéger les informations et les services contre les menaces externes et internes. Pour atteindre cet objectif, un système doit répondre à trois exigences fondamentales : la confidentialité, l'intégrité et la disponibilité [4]. La confidentialité garantit que seules les parties autorisées peuvent accéder aux données et aux services, tandis que l'intégrité garantit que seules les parties autorisées peuvent les modifier. La disponibilité garantit que les données et les services sont accessibles aux parties autorisées, même en cas d'attaque [5]. Ces propriétés sont mises en œuvre par le biais de politiques de sécurité, qui définissent les actions autorisées et interdites. La sécurité peut être abordée de manière **proactive**, grâce à des techniques telles que la programmation sécurisée, le renforcement de la configuration, l'authentification, le contrôle d'accès et la sécurité des communications. Elle peut également être abordée de manière **réactive** en détectant, en limitant et en corrigeant les menaces. La sécurité réactive implique de maintenir des représentations précises et des historiques d'activité des composants du système, souvent grâce à des outils de suivi d'activités et d'analyse distribuée. Les mesures de réactions et défense et autonomes telles que les logiciels antivirus et les outils de surveillance aident à détecter les abus et les anomalies dans le comportement du système, ce qui permet aux opérateurs de comprendre et de répondre aux menaces de manière efficace. Pour cela, ils doivent non seulement comprendre l'architecture de leur environnement, mais également les événements (normaux et anormaux) qui peuvent y survenir pour pouvoir ensuite émettre des hypothèses sur le comportement futur du système. Ce concept de Situational Awareness est adopté par la cybersécurité afin d'offrir des outils informatiques adaptés aux processus mentaux des agents. Les opérateurs chargés de la cybersécurité réactive sont généralement regroupés au sein de Centres d'Opérations Cyber (en anglais : Security Operation Centers ou SOC), qui sont généralement des lieux physiques, où différents opérateurs humains ont pour mission de détecter les intrusions

et y répondre. Pour être capable de répondre le plus rapidement possible et correctement aux alertes, ces opérateurs ont besoin de comprendre ce qu'il se passe sur les systèmes qu'ils surveillent. Cependant, même avec une bonne conscience de situation et des outils efficaces, certaines menaces restent difficiles à détecter, car noyées dans le « bruit » du comportement normal du système, ce type de menace correspond à la définition de signaux « faibles ». Dans ce chapitre, nous commençons par définir le concept de Situational Awarenness pour expliquer comment il a été adapté au domaine de la cybersécurité. Puis, nous décrivons le fonctionnement des SOCs, en particulier les rôles des opérateurs, les outils qu'ils utilisent et les processus qu'ils suivent. Enfin, nous montrons que les activités des opérateurs correspondent à rechercher des signaux « faibles » dans de grands volumes de données.

1.1 Cyber Situational Awareness

1.1.1 Situational Awareness

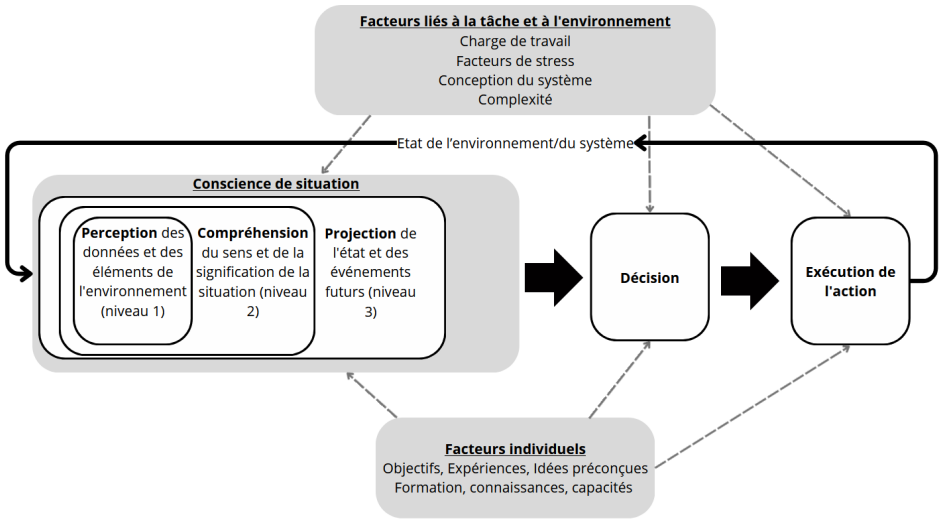


FIGURE 1.1 – Model de Situational Awareness d'Endlsey [6].

La Situational Awareness (SA) est la capacité cognitive des individus à analyser leur environnement et à réagir en conséquence, cette capacité joue un rôle essentiel dans la prise de décision. La SA comporte deux aspects distincts : l'aspect **technique**, qui implique la collecte, le traitement et la fusion de données pour en tirer des informations significa-

tives, et l'aspect **cognitif**, qui implique l'assimilation des données par un individu afin de prendre des décisions éclairées. Le modèle d'Endsley (Figure 1.1) identifie trois niveaux de Situational Awereness : la perception, la compréhension et la projection. Ces niveaux sont enveloppés les uns dans les autres plutôt que successifs, c'est-à-dire qu'une meilleure compréhension permet d'avoir une meilleure perception, et qu'une meilleure projection permet d'améliorer les deux niveaux précédents.

1.1.2 Cyber Situational Awareness

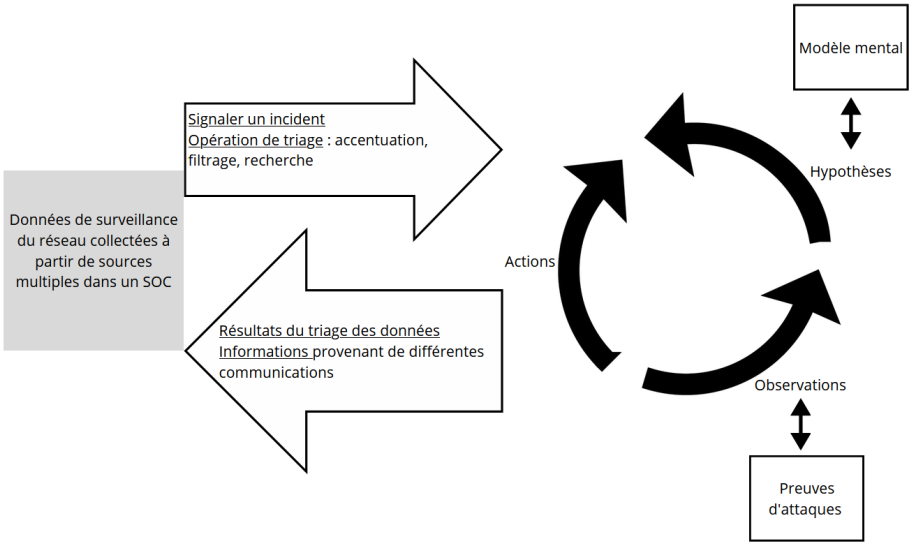


FIGURE 1.2 – Modèle de Cyber Situational Awareness de Zhong et al. [7]. Il sépare l'aspect cognitif de l'aspect technologique du traitement des informations de cybersécurité.

La Cyber Situational Awareness (CSA) peut être considérée comme une application de la Situational Awareness dans le domaine de la cybersécurité. La CSA nécessite une connaissance approfondie des activités de cybersécurité actuelles et passées d'une organisation afin de détecter et d'identifier efficacement les menaces et d'y répondre [7]. En se basant sur diverses sources, qui peuvent être techniques comme des scans bas niveaux ou plus abstraites comme des enjeux politiques, la CSA aide les opérateurs à comprendre les situations actuelle et future de leurs systèmes en matière de risques et de vulnérabilités ainsi que leurs capacités de protection et de réponse.

La CSA reprend les trois niveaux de la SA décrits précédemment et les adapte : perception (reconnaître les actifs et les situations de menace), compréhension (comprendre

la signification et l'impact des menaces), et projection (anticiper les menaces ou les actions futures) comme illustré par la figure 1.2.

1.1.3 Common Operationnal Picture

Les opérateurs chargés de la cybersécurité des organisations sont toujours organisés en équipes de tailles diverses (pouvant aller de quelques personnes à quelques dizaines de personnes). Chaque opérateur possède une conscience de situation cyber qui lui est propre et il est essentiel qu'il partage cette conscience avec les autres membres de l'équipe pour assurer le bon fonctionnement de celle-ci. Ainsi, les opérateurs collaborent afin de constituer la « Cyber Situational Awareness » de l'équipe. Pour l'instant, il n'existe aucun modèle de collaboration pour représenter la connaissance de situation en cybersécurité (CSA : Cyber Situational Awereness), qui est souvent examinée au niveau des individus. En effet, il est difficile d'évaluer la Situational Awareness d'une équipe et celle-ci est souvent vue comme la somme des Situational Awareness des personnes qui la composent [8]. Néanmoins, en s'inspirant de ce qui se fait dans l'armée, les Common Operational Pictures (COP), Esteve et al. [9] proposent une Cyber Common Operational Picture, qui prend en compte les niveaux stratégique, opérationnel et tactique/technique dans la conception des systèmes de représentations de données dédiés aux preneurs de décisions. Ils soulignent l'importance du couplage des capacités humaines avec celles des outils technologiques afin d'avoir un partage de la Situational Awareness entre les différents membres de l'équipe.

1.2 SOCs

Les équipes de cybersécurité réactives sont organisées en centres opérationnels de sécurité (SOC), où elles collaborent et utilisent leur connaissance collective de la situation cybersécurité (CSA) pour protéger les systèmes contre les attaquants. Les SOC jouent un rôle crucial au sein de l'organisation de défense d'une entreprise, en travaillant en parallèle avec les défenses proactives. Cependant, c'est au sein du SOC que l'optimisation de la CSA est primordiale, car les décisions prises en temps réel par l'équipe peuvent avoir un impact significatif sur le fonctionnement des systèmes protégés. Malgré des différences en termes de taille (de quelques opérateurs à quelques dizaines) et de moyens (d'une PME à une multinationale), la plupart des SOCs ont des fonctionnements très similaires en termes d'outils, de personnels, et de processus [10].

1.2.1 Personnels

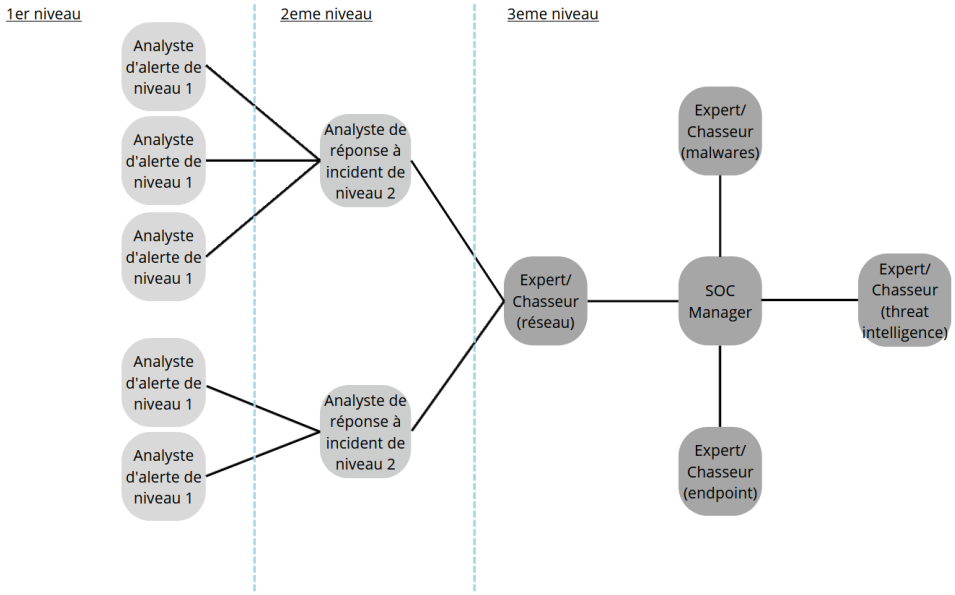


FIGURE 1.3 – Description des postes au sein des SOCs d'Alicia Torres [11]. Les opérateurs du premier niveau sont chargés de vérifier et de trier les alertes, tandis que ceux en deuxième et troisième niveau réalisent des analyses approfondies des situations.

Un SOC est principalement composé d'opérateurs (Figure 1.3) qui cherchent à protéger le ou les systèmes dont ils ont la charge [12]. Les opérateurs de premier niveau sont chargés de répondre aux alertes émises par les outils de sécurité dans la minute qui suit leur émission, en s'appuyant sur leur connaissance du système et des différents outils disponibles. Le traitement d'une alerte suit un processus bien défini (Figure 1.4), l'opérateur :

— reçoit l'alerte,
— prend connaissance de l'état du système,
— recherche des informations pour comprendre les causes de l'alerte,
— classifie l'alerte (FP, VP résolue, ne sait pas : on dit alors qu'il « escalade » l'alerte en la remontant vers les opérateurs de second niveau),
— rédige un rapport d'incident.

Si l'opérateur de premier niveau n'est pas capable de catégoriser l'alerte, il fait appel à un opérateur du deuxième niveau en escaladant l'alerte. Celui-ci est plus expérimenté que le premier et dispose généralement de davantage de droits sur le système comme le

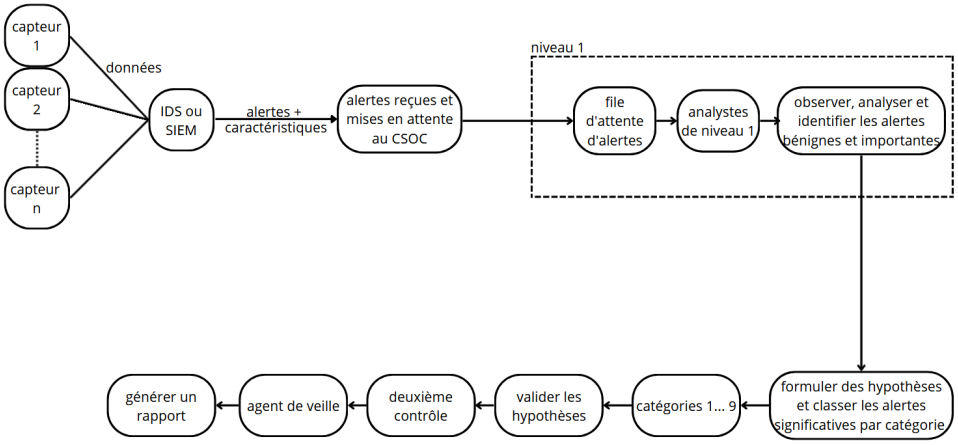


FIGURE 1.4 – Les alertes sont centralisées par le SIEM puis traitées individuellement par les opérateurs. [13].

changement des règles des pare-feu ou la modification de l'état du réseau. Les opérateurs de troisième niveau, qui sont souvent des opérateurs seniors, recherchent de manière proactive les menaces et les moyens d'améliorer la sécurité, et ont le droit de modifier le réseau, par exemple en modifiant son architecture. Dans les grandes entreprises, ils peuvent faire partie d'une équipe séparée qui collabore avec le SOC. Enfin, le responsable du SOC distribue les tâches et communique avec les autres équipes de cybersécurité et les échelons supérieurs.

Lorsqu'un opérateur escalade une alerte, il ajoute également des informations à propos du contexte de celle-ci (adresses IP, utilisateur suspect...) pour permettre à l'opérateur suivant de comprendre et de réagir plus rapidement à l'alerte. Une fois l'alerte traitée, si c'est un Vrai Positif, l'opérateur rédige un rapport d'incident afin d'une part de quantifier l'apport des SOCs à l'entreprise, mais aussi d'autre part pour savoir comment réagir si à l'avenir un problème similaire apparaît.

Physiquement, les SOCs ont souvent la même organisation spatiale (Figure 1.5) afin de faciliter la collaboration et le partage d'informations. Un grand écran permet de représenter les métriques générales du système surveillé. Les opérateurs du premier niveau sont les plus proches de l'écran tandis que les opérateurs plus expérimentés ainsi que les managers sont plus proches du fond. Les SOCs disposent parfois également d'une pièce annexe afin de permettre aux opérateurs de niveau 3 habilités de se rencontrer pour résoudre un problème particulièrement difficile.

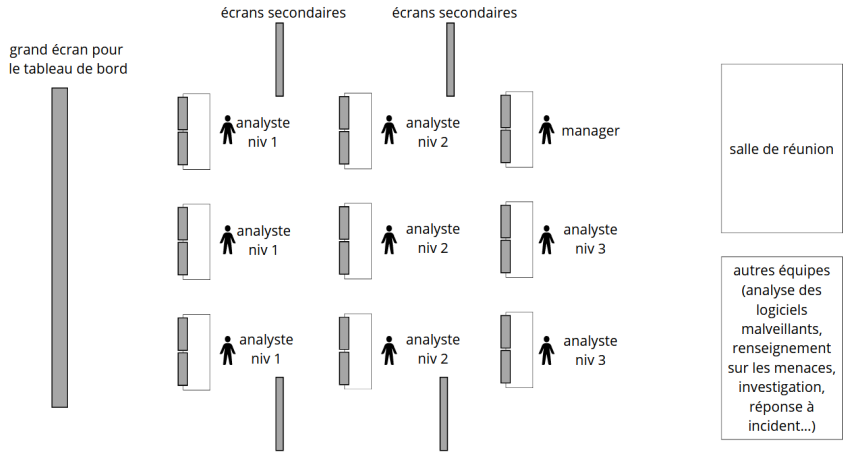


FIGURE 1.5 – Les opérateurs de niveau 1 se positionnent près d'un écran géant, tandis que les opérateurs de niveaux 2 et 3, ainsi que le responsable du SOC, prennent une position à l'arrière [12].

1.2.2 Outils

Afin de surveiller le réseau et de catégoriser les alertes, les agents des SOCs sont aidés par des systèmes de détection d'intrusion (Intrusion Detection System, IDS) [14]. Ceux-ci viennent compléter des dispositifs de défense de base comme les pare-feu et les antivirus. En surveillant le comportement des utilisateurs ou en inspectant les paquets qui entrent et sortent du réseau de l'organisation, ils recherchent des signes d'activités malveillantes, et alertent les agents lorsqu'une éventuelle tentative d'intrusion est détectée.

Il existe deux grandes catégories de détection d'intrusion : la détection basée sur des règles et la détection d'anomalie [15]. Dans la première catégorie, l'IDS détecte des modèles d'attaques connues, dans la seconde des profils avec un comportement « normal » sont créés et tout ce qui s'écarte de la norme provoque une alerte. Les systèmes basés sur la première catégorie détectent les attaques avec précision, mais uniquement pour des signatures connues, et sont inefficaces contre des attaques nouvelles ou complexes. Les systèmes basés sur la seconde catégorie (la détection d'anomalie) quant à eux sont capables de détecter de nouvelles attaques, mais leur efficacité dépend des données sur lesquelles ils ont été entraînés, car ils ne peuvent apprendre que les caractéristiques des comportements des systèmes sur lesquels ils ont été entraînés. Ces deux types de détection d'intrusion ont le gros inconvénient de créer beaucoup de faux positifs que les opérateurs doivent traiter. Afin de réduire le nombre d'alertes, ils passent beaucoup de temps à régler

les seuils d'alertes de ces dispositifs pour les rendre utilisables en limitant le nombre de faux positifs [16].

Une fois une alerte levée, les outils de détection ne fournissent pas forcément d'aide à la recherche et à l'inspection des données suspectes. Certains systèmes peuvent fournir des indications comme l'endroit où l'alerte a lieu. Les opérateurs doivent alors utiliser des outils d'analyse de trafic réseau plus ou moins sophistiqués comme Wireshark et Snort (Figure 3). Wireshark permet par exemple d'analyser les paquets de données échangés sur un réseau et Snort analyse en temps réel les paquets circulant sur un réseau et est capable de lever des alertes lorsqu'il détecte un comportement suspect. Pour afficher les résultats de ces outils, les opérateurs utilisent des tableurs comme Excel avec plusieurs fenêtres et tableaux croisés dynamiques afin de gérer et de corréler le grand nombre de données qu'ils doivent analyser. Les opérateurs utilisent ces outils pour effectuer des requêtes successives pour exclure manuellement les données inutiles. En effet, ils ne souhaitent pas utiliser de filtres qui risqueraient de supprimer des éléments à conserver [17].

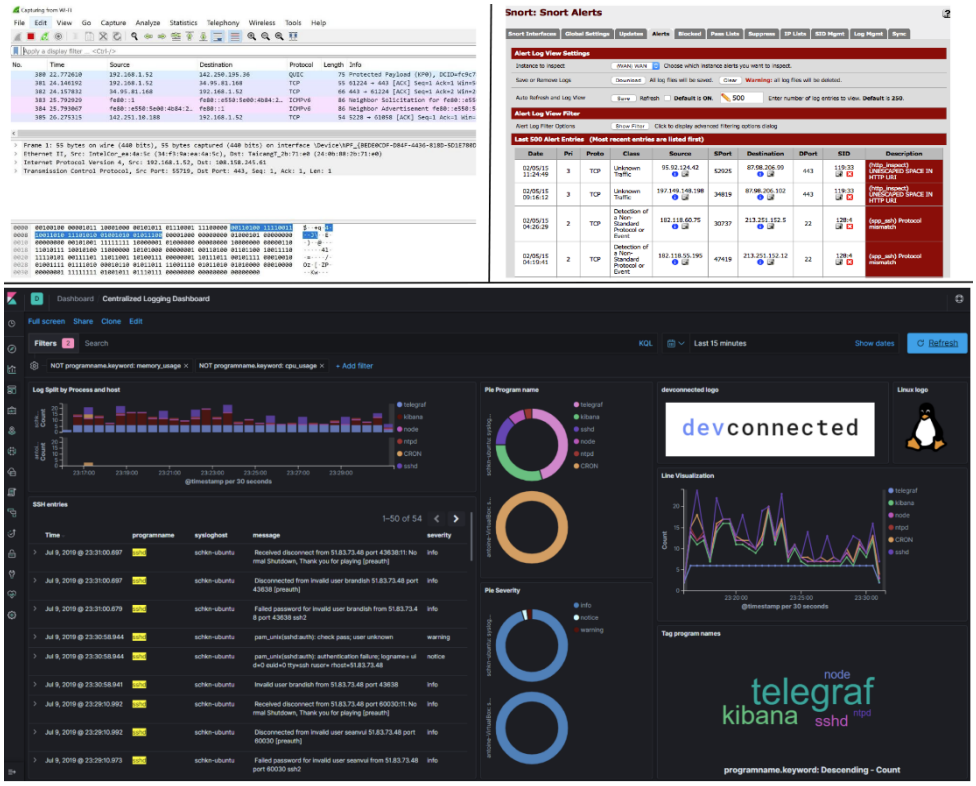


FIGURE 1.6 – Exemple d'interfaces d'outils de la cybersécurité. En haut à gauche Wireshark, en haut à droite Snort, et en bas Kibana.

A partir d'un certain niveau de maturité, les SOCs utilisent un outil de gestion des informations et des événements de sécurité (SIEM pour Security Information and Event Management) pour agréger les données et les alertes levées par les différents capteurs pour faciliter la recherche d'anomalies et la corrélation des données. Depuis quelques années, des solutions « clés en main » comme Splunk ou la Suite Elasticsearch (Figure 1.6) permettent d'agréger les données et alertes des différents capteurs afin de faciliter la réponse aux alertes. Ces outils permettent d'effectuer des requêtes avec des langages ressemblant au SQL dont les réponses s'affichent sous format de tableaux. Il est également possible de présenter des graphiques simples et interactifs afin de contextualiser les données et de les filtrer plus facilement, comme sélectionner une plage de temps ou un type d'alerte particulier. Cependant, les visualisations sont peu utilisées par les opérateurs réseaux qui préfèrent utiliser les outils tabulaires et des successions de requêtes. Ils trouvent que les visualisations « lissent » trop les données [18], c'est-à-dire qu'elles ne permettent pas de voir assez en détail les données, ce qui empêche la détection des anomalies.

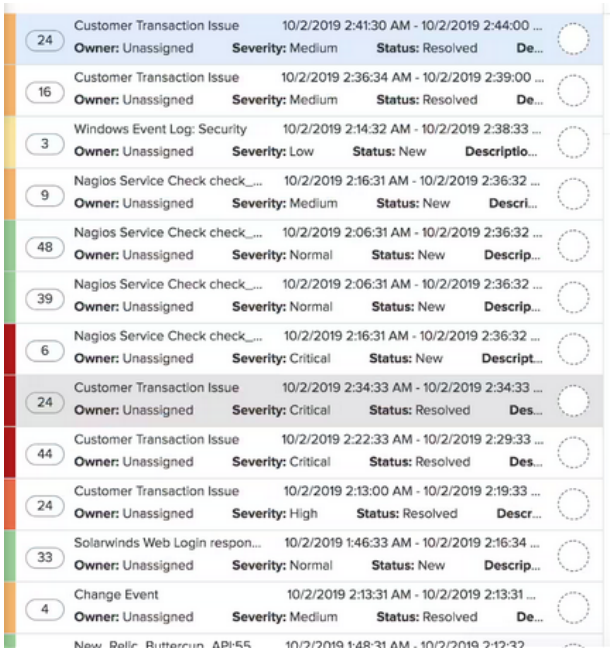


FIGURE 1.7 – Interface de ticketing d'un SIEM

Un outil central des SOCs est leur système de ticketing (Figure 1.7). Les alertes levées par les SIEMs sont fournies aux opérateurs par une liste de tickets, l'opérateur choisit alors les tickets les plus importants pour les gérer en premier. Ce système de ticketing peut permettre de transmettre des informations sur l'alerte entre les opérateurs. Cependant, les

opérateurs vont souvent utiliser d'autres moyens de communications comme un document partagé ou tout simplement l'oral. Cette communication peut parfois être difficile en fonction de la charge de travail ainsi que de l'organisation du SOC. C'est d'autant plus vrai que certains SOCs sont maintenant distribués à travers le monde, ce qui leur permet d'être opérationnels 24h/24 au détriment de la communication entre les opérateurs.

1.3 Traitement des alertes et signaux faibles

Ansoff [19] a introduit le concept de signaux faibles en 1975 comme un aspect essentiel de l'analyse de l'environnement et de l'identification des opportunités stratégiques pour les entreprises. Les signaux faibles sont des indications précoces, subtiles ou faibles, de changements significatifs ou de tendances émergentes susceptibles d'avoir un impact sur l'environnement ou le secteur d'activité d'une organisation. Ces signaux sont souvent négligés ou ignorés en raison de leur manque de visibilité immédiate ou de reconnaissance généralisée. Cependant, ils offrent des indications précieuses sur les perturbations ou les changements potentiels dans l'environnement de l'entreprise. Ansoff a insisté sur la nécessité d'une analyse et d'une surveillance actives de l'environnement externe afin d'identifier les signaux faibles. Les organisations doivent mettre en place des mécanismes et des processus permettant de capter, d'interpréter et d'analyser efficacement ces signaux. Ce faisant, elles peuvent prendre des décisions en connaissance de cause et adapter leurs stratégies pour tirer parti des tendances émergentes ou atténuer les risques.

D'autres chercheurs, tels que Lesca [20], Coffman [1] et Sidhom [21], ont donné une définition des signaux faibles dans le contexte des ressources de santé. Ils décrivent les signaux faibles comme des informations fragmentaires noyées dans une masse de bruit, présentant une ambiguïté apparente, une faible utilisabilité et un manque de « palpabilité ».

Le concept de signaux faibles existe aussi en criminalité, selon le Service canadien de renseignements criminels (SCRC) [1], des indicateurs isolés peuvent ne pas avoir de valeur tangible et être symptomatiques de divers phénomènes sans rapport avec la criminalité organisée. Toutefois, lorsque ces indicateurs sont regroupés en fonction de conditions telles que la proximité temporelle ou géographique d'activités spécifiques, ils peuvent donner un aperçu de la présence ou de l'émergence d'un comportement criminel. C'est donc la corrélation de plusieurs indicateurs qui indique un danger [22].

1. https ://legacy.mgtaylor.com/mgtaylor/jotm/winter97/wsrintro.htm

En nous basant sur les définitions précédentes, nous proposons cette définition de signaux faibles : *Avertissements difficiles à détecter, trop incomplets pour permettre des estimations précises, qui nécessitent de corréler plusieurs indicateurs pour être utilisables.* Un indicateur important est la périodicité des évènements du système, à la fois pour la baseline et pour les malwares.

Les évènements apparaissant dans un système d'entreprises sont généralement réguliers et prévisibles [23]. Un système a une « vie » propre que l'opérateur apprend à connaître, ce qui lui permet ainsi de détecter différents comportements suspects, car déviant de la norme. Une stratégie fréquemment employée par les opérateurs consiste à identifier d'abord ce qui est normal, puis à utiliser cette définition pour mettre en évidence ce qui est anormal [24]. De plus, à chaque phase de l'attaque (accès initial, première pénétration, élargissement de l'accès, mise en place de l'attaque, exfiltration ou dégâts), l'attaquant laisse des traces périodiques détectables par des logiciels de détections d'anomalies spécialisés dans l'analyse de signaux temporels [25]. Toutefois, il existe un large éventail d'applications légitimes qui ont également un comportement périodique, par exemple pour faire des mises à jour. Par conséquent, la détection de signaux périodiques dans le trafic réseau pour les malwares est vouée à générer de nombreuses fausses alertes positives, ce qui exacerbe le problème de la fatigue des alertes auxquelles sont confrontés les opérateurs [26].

Résumé

Les entreprises s'appuient sur les opérateurs des centres opérationnels de sécurité (SOC) pour protéger leurs systèmes informatiques. Les SOCS fournissent des outils informatiques qui permettent aux opérateurs de comprendre et prédire l'état du système qu'ils surveillent. Les opérateurs utilisent les alertes générées par ces outils comme point de départ pour comprendre les menaces qui pèsent sur le système. Ils utilisent ensuite leur connaissance du comportement nominal du système pour analyser les données associées aux alertes. Cependant, l'approche actuelle, qui consiste à utiliser des outils tabulaires pour explorer la quantité écrasante de données et d'alertes, peut s'avérer fastidieuse et chronophage. Bien que des visualisations puissent faciliter ce processus, leur adoption dans les environnements SOC est actuellement limitée.

Chapitre 2

Visualisations pour la cybersécurité

Dans le chapitre précédent, nous avons expliqué comment les outils de cybersécurité des SOCs sont efficaces, mais saturés en raison de la trop grande quantité de données qu'ils doivent traiter. Pour résoudre ce problème et améliorer la conscience de la situation, la communauté de la visualisation scientifique propose des visualisations simples, mais qui ne sont pas encore adoptées par les professionnels. Dans ce chapitre, nous nous intéressons à la visualisation des données et à son application dans le domaine de la cybersécurité, ainsi qu'à ses avantages et à ses limites. Nous commencerons par présenter le domaine de la visualisation des données, en particulier le cadre de Munzner [3] qui facilite la description des systèmes de visualisation en fonction de divers paramètres. En outre, nous présenterons des directives pour obtenir des représentations visuelles efficaces et discutons brièvement de l'intérêt de la 3D pour les visualisations. Ensuite, nous utiliserons le cadre de Munzner pour décrire les visualisations de cybersécurité utilisées pour répondre aux alertes. Grâce à cela, nous identifierons les tendances des techniques de visualisations utilisées pour la cybersécurité, nous mettrons en évidence leurs avantages et leurs inconvénients, et nous comparerons leurs propositions avec les habitudes de travail des opérateurs.

2.1 Visualisations

2.1.1 Présentation

Au cours des dernières décennies, la visualisation s'est imposée comme une discipline scientifique ayant de profondes implications dans divers domaines. Les techniques de visualisation ont joué un rôle essentiel dans la transformation de données complexes en représentations intuitives et informatives, permettant aux humains de comprendre et d'analyser les données de manière plus efficace. Les progrès dans le domaine des technologies

numériques, incluant tant la puissance de calcul que la qualité des écrans (en termes de couleur et de résolution), ont facilité l'affichage de volumes importants d'informations. De plus, ces avancées ont permis d'ajuster en temps réel les visualisations en réponse aux demandes des utilisateurs, grâce à l'intégration de nouvelles formes d'interactions. C'est dans les années 80 que les chercheurs s'emparent de la question de la visualisation assistée par ordinateur, de cette époque datent plusieurs définitions des visualisations :

— McCormick et al. (1987) [27] formulent l'idée clé de la visualisation comme suit : « La visualisation est une méthode de traitement informatique. Elle transforme le symbolisme en géométrie, ce qui permet aux chercheurs d'observer leurs simulations et leurs calculs. La visualisation offre une méthode pour voir l'invisible. Elle enrichit le processus de découverte scientifique et favorise des prises de conscience profondes et inattendues ».

— Card et al. (1999) [28] définissent la visualisation, celle de l'information en particulier, comme suit : « L'utilisation de représentations visuelles interactives des données, assistées par ordinateur, pour amplifier la cognition ». Au cœur de cette définition, les capacités de perception et de cognition de l'homme et les capacités de calcul des ordinateurs sont considérées comme des éléments clés liés par l'interaction.

— Ware (2008) [29] insiste sur l'interaction entre l'homme et l'ordinateur : « Il est utile de considérer l'homme et l'ordinateur comme une seule entité cognitive, l'ordinateur fonctionnant comme une sorte de coprocesseur cognitif du cerveau humain. Chaque partie du système fait ce qu'elle fait le mieux. L'ordinateur peut pré-traiter de grandes quantités d'informations. L'homme peut effectuer une analyse rapide des schémas et prendre des décisions souples ».

Les différentes définitions font bien le lien entre cognition et informatique, ce qui est particulièrement important pour la CSA. Pour autant, pour que l'humain et la machine fonctionnent en tant qu' « entité cognitive », il faut des visualisations efficaces à la fois dans le traitement des données, mais aussi dans la façon de les transmettre à l'humain. Ainsi, certains définissent des critères essentiels pour transformer les données en représentations visuelles. Par exemple, Van Wijk (2006) [30] souligne l'importance de s'assurer que les avantages de la visualisation l'emportent sur les coûts qui y sont associés, notamment les efforts de calcul et d'interprétation. Pour obtenir des représentations visuelles efficaces et bénéfiques, il est essentiel de prendre en compte deux critères fondamentaux identifiés par Mackinlay (1986) [31] : l'**expressivité** et l'**efficacité**. L'expressivité concerne la capacité

des représentations visuelles à transmettre l'information voulue avec précision, tandis que l'efficacité concerne l'optimisation de l'utilisation du système visuel humain et des capacités du support de sortie.

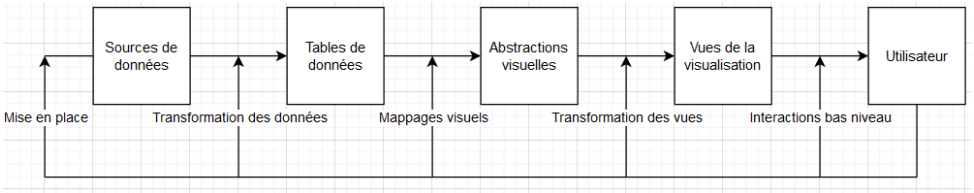


FIGURE 2.1 – Le pipeline de la visualisation

Pour répondre aux critères d'expressivité et d'efficacité, les approches de visualisation suivent souvent le modèle du « pipeline de visualisation » (Figure 2.1). Ce modèle sert non seulement de cadre à la description et à la mise en œuvre du processus de génération de représentations visuelles des données, mais il permet également de définir le rôle de l'utilisateur dans ce processus. Card et al. (1999) ont présenté le pipeline de visualisation largement accepté, qui décrit les étapes successives par lesquelles les données sont transformées de leur source en tableaux de données, en abstractions visuelles et finalement en vue(s) de visualisation. Les premières étapes du pipeline englobent diverses transformations de données, telles que le filtrage, le regroupement et la correction des erreurs. L'élément central du pipeline repose sur les mappages visuels, qui consistent à transformer les tableaux de données en abstractions visuelles afin de créer des visualisations pertinentes. Les nombreuses recherches qui permettent de choisir la vue la plus pertinente en fonction des données seront présentées par la suite.

Dans le contexte du pipeline de visualisation, l'utilisateur joue un double rôle. Premièrement, en tant que destinataire des informations véhiculées par les représentations visuelles, l'objectif de l'utilisateur est de comprendre la signification des graphiques et d'établir des liens entre les éléments visuels et les attributs correspondants des données. Deuxièmement, ce modèle propose que l'utilisateur puisse superviser et guider les différentes étapes de la transformation des données brutes vers la visualisation finale.

2.1.2 Catégorisation de visualisations

Pour décrire les systèmes de visualisations, nous avons choisi le cadre théorique de Munzner [3] car c'est le seul qui permette de prendre en compte le but des utilisateurs

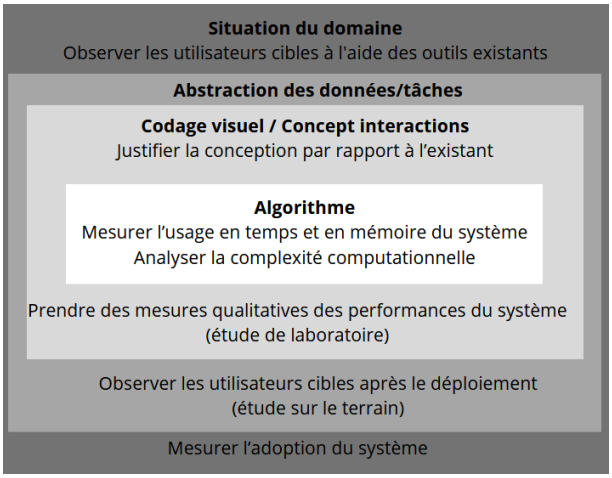


FIGURE 2.2 – Les quatre niveaux d'évaluations du cadre de Munzner [3].

dans la description et le choix des visualisations [32]. En effet, il permet de relier les concepts de visualisations de bas niveaux comme les types de visualisations et d'interactions utilisés avec le contexte d'utilisation de la visualisation grâce à une description des tâches de l'utilisateur (son but et sa cible en utilisant la visualisation). Le cadre proposé par Munzner prend en compte le contexte d'utilisation de la visualisation, l'abstraction des données (le **quoi**) et des tâches (le **pourquoi**), l'encodage visuel et les interactions (le **comment**) de la visualisation, ainsi que l'algorithme qui a créé la visualisation. Il est possible d'évaluer les performances des visualisations à tous les niveaux (Figure 2.2). Ce cadre peut être utilisé pour décrire les visualisations de cybersécurité afin de les exprimer dans un langage compréhensible pour les experts en visualisation. Nous nous concentrons sur le **quoi**, le **pourquoi** et le **comment** de la visualisation, car le contexte d'usage ne peut pas être décrit par un cadre de visualisation, mais par son langage métier, et parce que les visualisations informatiques dépendent de nombreuses librairies techniques différentes pouvant être utilisées. Ensemble, ces trois éléments constituent une « instance » d'analyse (Figure 2.3). Pour un outil de visualisation de base, une seule instance suffirait, mais pour des outils plus complexes, une série d'instances interconnectées est nécessaire pour les décrire complètement. Nous décrivons en détail les trois étapes de l'analyse les sections suivantes.

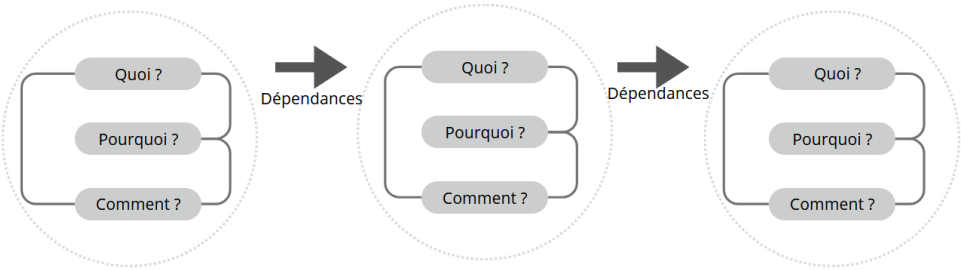


FIGURE 2.3 – Le cadre de Munzner permet d'analyser les visualisations comme des séquences enchaînées d'instances, où la sortie d'une instance est l'entrée d'une autre. Chaque instance peut être décrite par le type de données visualisées (what), l'intention de l'utilisateur (why), et les techniques de visualisation utilisées (how).

What

Selon la classification de Bertin [33], l'organisation des données pour l'affichage peut être classée en trois niveaux : qualitatif (nominal), ordonné et quantitatif (métrique). Au niveau qualitatif, les concepts présentent une différenciation simple et les catégories ne sont pas universellement ordonnées, ce qui offre une certaine souplesse pour différents arrangements. Le niveau ordonné, quant à lui, comprend des catégories qui peuvent être organisées de manière universelle et qui sont équidistantes les unes des autres, ce qui permet d'établir un ordre cohérent. Enfin, le niveau quantitatif implique des valeurs numériques qui indiquent la variation de la distance entre les catégories, donnant un sens aux distances et permettant la création de groupes distincts sur la base de ces mesures. La théorie des échelles de mesure proposée par Stevens [34] reprend celle de Bertin mais sépare les données quantitatives en Intervalle et Ratio. Les données d'intervalle possèdent les propriétés des échelles nominales et ordinales et possèdent un intervalle régulier entre deux valeurs. Les ratios englobent toutes les propriétés des échelles précédentes et utilisent en outre la valeur nulle pour des comparaisons et des opérations mathématiques plus complètes. L'approche de Ware [35] s'aligne sur celle de Bertin et Steven en définissant deux formes fondamentales pour les données : la valeur des données (entité) et la structure des données (relation). Les entités comme les relations peuvent être décrites à l'aide de l'échelle de Stevens. Le cadre de Munzner [3] classe les informations utilisées dans la visualisation des données en cinq types fondamentaux : les éléments, les attributs, les liens, les positions et les grilles. Les attributs font référence à des propriétés spécifiques mesurables telles que le salaire, le prix ou la température. Les éléments représentent des entités individuelles, discrètes et distinctes, telles que des personnes, des actions ou des

villes. Les liens sont les connexions entre les éléments, qui forment souvent des réseaux ou des arbres. Les grilles définissent la stratégie d'échantillonnage des données continues sur la base des relations géométriques et topologiques entre les cellules. Les positions fournissent des données spatiales, spécifiant des emplacements dans l'espace 2D ou 3D, comme les paires latitude-longitude ou les coordonnées dans les scanners médicaux.

Comme le montre Ware, la valeur des attributs est importante, mais ne suffit pas à décrire les données, leur organisation dans le jeu de données est également extrêmement importante. Le cadre de Munzner [3] classe les ensembles de données en quatre types principaux : les tableaux, les réseaux, les champs et la géométrie. Les tableaux comportent des cellules indexées par des éléments et des attributs, soit sous forme simple et plate, soit sous forme complexe et multidimensionnelle. Les réseaux sont constitués de nœuds reliés par des liens, les arbres constituant un cas particulier. Les champs continus utilisent des grilles dont les positions spatiales contiennent des attributs. La géométrie spatiale implique uniquement des informations sur la position.

Les données ordonnées peuvent être classées en deux catégories : les données séquentielles et les données divergentes. Les données séquentielles présentent une plage homogène allant d'une valeur minimale à une valeur maximale, tandis que les données divergentes peuvent être décomposées en deux séquences pointant dans des directions opposées, convergeant vers un point zéro commun. En outre, les données ordonnées peuvent être cycliques, c'est-à-dire que les valeurs reviennent à un point de départ au lieu d'augmenter indéfiniment. Les mesures temporelles, telles que l'heure du jour, le jour de la semaine et le mois de l'année, présentent souvent un comportement cyclique.

Dans la visualisation des données, un attribut clé joue un rôle essentiel en tant qu'index permettant de rechercher des attributs de valeur. Pour les tables à dimensions multiples, une clé unique peut être obtenue par une combinaison de champs. Les attributs temporels peuvent avoir une sémantique de valeur ou de clé. La durée du temps écoulé ou la date d'une transaction sont des exemples d'attributs temporels ayant une sémantique de valeur dépendante. Dans les champs spatiaux et les tables abstraites, le temps peut servir de clé indépendante.

Les séries de données temporelles sont un cas courant de données temporelles, consistant en une séquence ordonnée de paires temps-valeur. Ces ensembles de données sont un cas particulier de tables, où le temps sert de clé. Les paires de valeurs temporelles sont souvent, mais pas toujours, espacées d'intervalles temporels uniformes. L'analyse des données de séries temporelles implique des tâches telles que l'identification des tendances,

des corrélations et des variations à plusieurs échelles de temps, telles que les fréquences horaires, quotidiennes, hebdomadaires et saisonnières.

Pourquoi

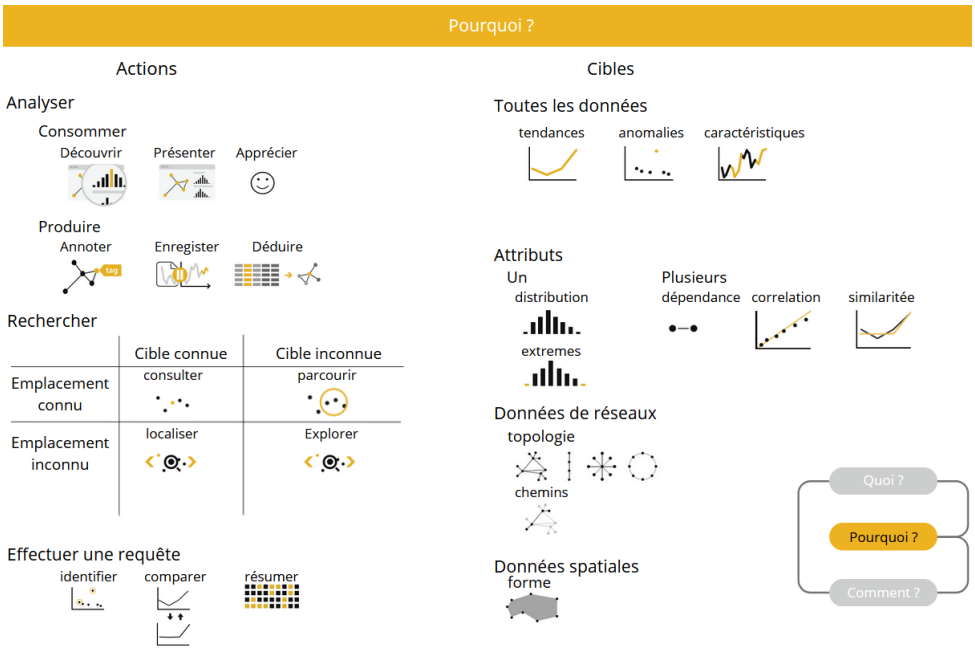


FIGURE 2.4 – La partie « pourquoi » du cadre de Munzner [3] (2014, ©Taylor and Francis Group LLC) permet d'expliquer pourquoi les utilisateurs utilisent la visualisation en termes d'actions et de cibles.

Les types de visualisations utilisés sont certes influencés par les types de données, mais aussi par le contexte d'utilisation et les besoins de l'utilisateur. En effet, des intentions et différents contextes peuvent conduire à des visualisations distinctes, même lorsque l'on examine des données identiques. Cantu et al. [36] ont montré que les transformations des visualisations sont influencées par le contexte d'utilisation. La prise en compte de ce contexte et des tâches de l'opérateur est cruciale pour une visualisation efficace. Cette idée est reprise par Munzner pour décrire l'intention de l'utilisateur lorsqu'il utilise la visualisation, c'est la partie "why" du cadre (Figure 2.4). Cette partie se décompose en actions et cible les raisons pour lesquelles un outil de visualisation est utilisé. Les actions de plus haut niveau consistent à utiliser les visualisations pour utiliser ou produire de l'information. Les cas d'utilisation sont la présentation, la découverte et le plaisir ; la

découverte peut impliquer la génération ou la vérification d'une hypothèse. Au niveau intermédiaire, la recherche peut être classée selon la connaissance ou non de l'identité et de l'emplacement des cibles : les deux sont connus dans le cas de la recherche, la cible est connue, mais son emplacement ne l'est pas dans le cas de la localisation, l'emplacement est connu, mais la cible ne l'est pas dans le cas de la navigation, et ni la cible ni l'emplacement ne sont connus dans le cas de l'exploration. Au niveau le plus bas, les requêtes peuvent avoir trois buts : identifier une cible, comparer certaines cibles, et enfin présenter une vue d'ensemble de toutes les cibles. Pour toutes les données, l'objectif de l'utilisateur peut être de trouver des tendances, des aberrations ou des caractéristiques (souvent propres au contexte). Pour un attribut, la cible peut être une valeur, les extrêmes des valeurs minimales et maximales ou la distribution de toutes les valeurs sur l'ensemble de l'attribut. Pour plusieurs attributs, la cible peut être les dépendances, les corrélations ou les similitudes entre eux. Pour les données de réseau, la cible peut être la topologie en général ou les chemins en particulier, et pour les données spatiales, la cible peut être la forme.

How

Pour décrire les visualisations, Munzner utilise le concept de « mark » (éléments graphiques de base classés par dimensions spatiales) et de « channel » (moyen visuel de contrôler l'apparence des marques) pour décrire chaque élément précis de la visualisation. Ce type de typologie est très précis, mais, selon ses propres dire, rend l'analyse de nombreuses visualisations fastidieuses. D'autres typologies utilisent des descriptions plus générales pour décrire les visualisations connues, nous nous inspirons de l'une d'entre elles qui décrit 10 types de visualisations fréquemment utilisées en cybersécurité [37] : diagrammes nœud-lien, matrices, coordonnées parallèles, timelines (des représentations de données où l'axe horizontal représente le temps et l'axe vertical représente une variable quantitative, en cybersécurité, elles peuvent prendre la forme d'histogrammes, de graphiques linéaires, de diagrammes à horizon), treemaps, cartes géographiques, nuages de points, nuages de mots, tables, données brutes. Nous avons ajouté d'autres visualisations, car elles ne pouvaient pas entrer dans cette liste, mais elles demeurent minoritaires dans le corpus (cyclique, petri, heatmap, gantt, formes géométriques, liens, « sunburst », diagramme radar). Comme les visualisations 2D et 3D qui se ressemblent reposent sur le même concept, nous ne les différencions pas dans notre catégorisation (par exemple, un nuage de points 3D repose sur le même concept qu'un nuage de points 2D). La com-

paraison avec le comportement nominal étant importante, nous ajoutons une catégorie décrivant (voir section 2) la méthode de visualisation utilisée pour représenter le comportement temporel des données.

Nous utilisons le cadre de Munzner pour décrire comment les différentes vues sont coordonnées entre elles (par la sélection des données, par l'apparition de tooltips, en juxtaposant les informations visuelles, en superposant les informations visuelles, en créant des liens entre les informations), et comment les données sont visualisées différemment entre les vues (multiforme, multiforme avec overview+detail).

Pour décrire les interactions, nous utilisons le cadre de Brehmer et Munzner [38] qui présente 10 catégories : naviguer (glisser, saisir le monde, monde en miniature, zoomer), sélectionner (sélection d'un seul objet, sélection de plusieurs objets par brossage ou boîtes de sélection), détails sur demande (approches réactives, approches prédictives, écran 2D supplémentaire), arranger (points de données, composants, vues), modifier (mise en évidence, mappage d'attributs, représentations), filtrer (sélection directe, couche abstraite), agréger, annoter, importer, dériver, enregistrer. Nous classons également par catégorie la manière dont l'utilisateur peut interagir avec les outils de cybersécurité au sein du système de visualisation : IDS, console, système d'interrogation.

2.1.3 Bonnes pratiques

Pour améliorer la conception des outils de visualisation, il est essentiel de prendre en compte les contraintes des utilisateurs et de comprendre leurs objectifs. Bien que les tests et les évaluations soient essentiels pour évaluer le système, ils peuvent nécessiter beaucoup de temps et de ressources. Pour optimiser ce processus, la littérature existante sur la cognition et les capacités humaines peut fournir des informations sur les bonnes pratiques à adopter pour créer des systèmes de visualisation performants.

Perception humaine

Comme la visualisation se base sur la vision humaine, quelques règles sont importantes à connaître afin de réaliser des visualisations efficaces. Dans cette section, nous donnons un aperçu des différentes théories de la perception visuelle et explorons leurs applications dans le domaine de la visualisation.

Certains changements visuels peuvent capter notre attention plus efficacement que d'autres (Figure 2.5), la nature de ces changements étant étroitement liée aux propriétés

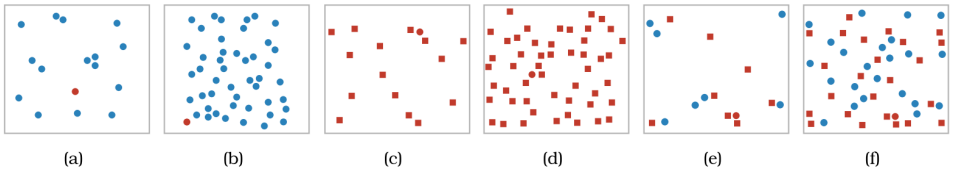


FIGURE 2.5 – Exemples du phénomène de « popout » [3] (2014, ©Taylor and Francis Group LLC). Le cercle rouge se détache d'un petit ensemble de cercles bleus. (b) Le cercle rouge se détache tout aussi rapidement d'un grand ensemble de cercles bleus. (c) Le cercle rouge se détache également d'un petit ensemble de formes carrées, bien que légèrement plus lentement qu'avec la couleur. (d) Le cercle rouge se détache également d'un grand ensemble de carrés rouges. (e) Le cercle rouge ne prend pas longtemps à être repéré dans un petit ensemble de formes et de couleurs variées. (f) Le cercle rouge ne se détache pas d'un grand ensemble de carrés rouges et de cercles bleus, et il ne peut être trouvé qu'en cherchant un par un à travers tous les objets.

visuelles des représentations utilisées. Par exemple, de petits changements de position ou de taille sont souvent plus perceptibles que des modifications graduelles de couleur ou de texture. La théorie de la perception pré attentive, proposée par Treisman [39], explique comment des propriétés visuelles spécifiques peuvent être perçues sans effort et rapidement (en moins de 250 ms). Par exemple, dans une image composée de cercles bleus, la perception immédiate d'un cercle rouge ne nécessite aucune charge cognitive et se produit automatiquement. Toutefois, ce processus peut s'interrompre si l'image contient un nombre excessif de couleurs différentes. Des propriétés telles que la couleur, l'orientation, la longueur, l'épaisseur, la courbure, etc. peuvent toutes contribuer à la perception pré-attentive [40]. Cependant, mixer certaines propriétés entre elles peut créer des interférences visuelles et nuire à la clarté de la visualisation [3] (Figure 2.6).

La théorie de la Gestalt [41] est centrée sur la synthèse mentale des formes en une « superforme » cohésive lors de la perception d'un objet (Figure 2.7). Plusieurs lois fondamentales régissent les principes de cette théorie. La loi de la bonne forme postule qu'un ensemble de parties suffisamment proches tend à être perçu comme une forme unifiée. La loi de proximité regroupe naturellement les objets proches les uns des autres, tandis que la loi de continuité suggère que les éléments lisses et continus forment des entités visuelles plus reconnaissables. De même, la loi de similarité indique que les éléments visuels ayant des propriétés communes ont tendance à se regrouper, et la loi du destin commun relie les parties mobiles ayant la même trajectoire comme faisant partie de la même forme. Enfin, la loi de fermeture suggère que les formes fermées sont plus facilement identifiées comme

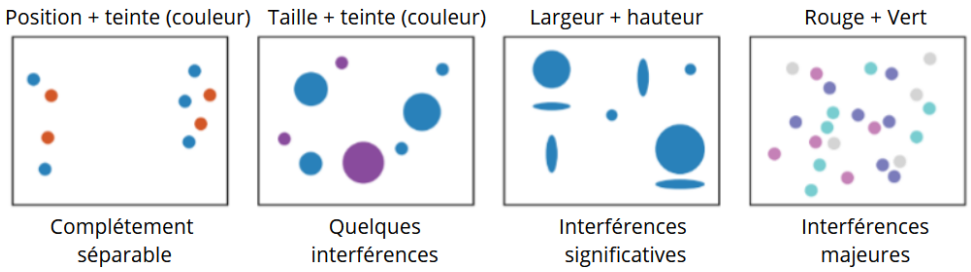


FIGURE 2.6 – Exemples d'interférences entre canaux visuels [3] (2014, ©Taylor and Francis Group LLC). La couleur et la localisation sont des canaux séparables bien adaptés pour coder différentes caractéristiques de données. Cependant, la taille interagit avec la teinte, ce qui est plus difficile à percevoir pour les petits objets. Les perceptions des distances horizontales et verticales fusionnent pour donner une perception de la surface, produisant ainsi trois canaux.

des formes que les formes ouvertes.

La visualisation des données utilise efficacement ces phénomènes de Gestalt pour permettre aux utilisateurs de discerner des motifs, des groupes et des valeurs aberrantes dans de vastes ensembles de données... Cependant, une mauvaise utilisation de ces principes peut conduire à l'identification de modèles dénués de sens, ce qui se traduit par des mauvaises interprétations des données. Pour éviter de tels cas, les systèmes de visualisations doivent faire l'objet de tests et être accompagnés d'un contexte suffisant pour écarter les motifs ambigus. En combinant les propriétés pré-attentives avec les principes de la Gestalt, les utilisateurs peuvent rapidement détecter et comprendre les corrélations ou les phénomènes émergents dans les données. Cette combinaison de techniques explique en partie pourquoi certaines variables visuelles et leurs combinaisons ont plus d'impact que d'autres dans la visualisation des données.

La loi de Weber (Figure 2.8), proposée par Weber et étudiée par Fechner [42], explore la perception d'un stimulus en fonction de son intensité. Elle démontre que la perception humaine est basée sur des jugements relatifs plutôt d'absolus. Ce principe s'applique à toutes les modalités sensorielles et a des implications importantes pour la visualisation. Par exemple, notre capacité à détecter les différences de longueur est un pourcentage de la longueur de l'objet. Cela a un impact sur la précision et la discriminabilité de nos perceptions. Les visualisations gagnent à prendre en compte la loi de Weber en fournissant des échelles et des alignements communs qui aident à porter des jugements plus précis. L'alignement des barres, par exemple, facilite les comparaisons et améliore la précision

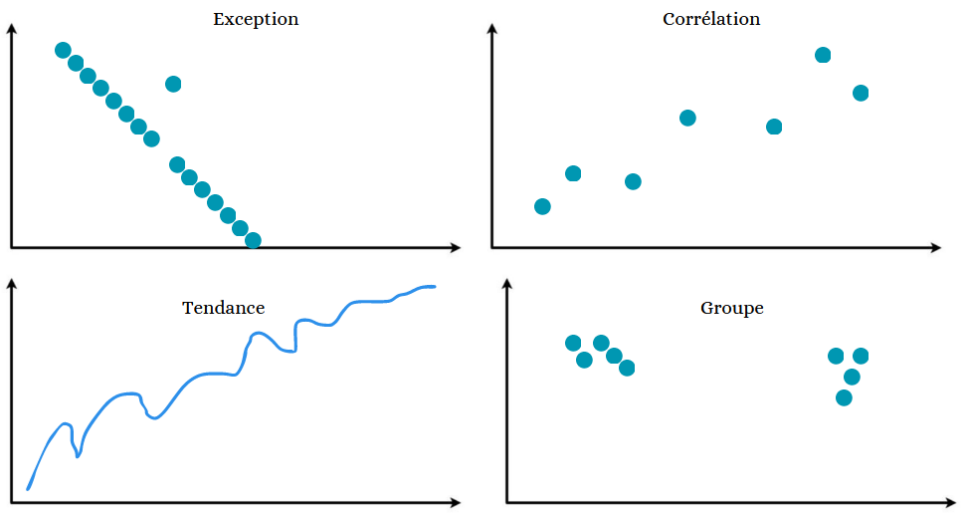


FIGURE 2.7 – Utilisation de la gestalt pour les visualisations de données.

des jugements de longueur.

La perception de la profondeur dans les visualisations 3D repose sur divers indices qui traduisent les relations spatiales et la structure tridimensionnelle (Figure 2.9). Les principaux indices comprennent l'occlusion, la distorsion de la perspective, les ombres et l'éclairage, qui jouent un rôle important en aidant les spectateurs à interpréter les distances relatives et les formes des objets dans la visualisation. Chacun de ces indices comporte des coûts et des inconvénients spécifiques. Par exemple, l'occlusion peut cacher des informations cruciales, tandis que la distorsion de la perspective peut avoir un impact sur la précision des mesures. En outre, l'utilisation d'ombres et d'ombres de surface contribue à la perception de la profondeur, mais nécessite des soins particuliers afin d'obtenir une représentation visuelle adéquate. Il est essentiel de comprendre l'interaction de ces indices de profondeur pour concevoir des visualisations 3D efficaces et significatives, car ils déterminent la manière dont les utilisateurs perçoivent et interprètent les relations spatiales dans les données représentées.

Guidelines

Une des premières métriques d'évaluation d'une visualisation est son ratio données représentées/encre utilisée (Data/ink ratio), qui évalue la proportion d'encre directement utilisée pour la représentation des données par rapport à l'encre totale utilisée. Maximiser ce rapport permet d'améliorer la lisibilité d'une visualisation. Un autre facteur critique est

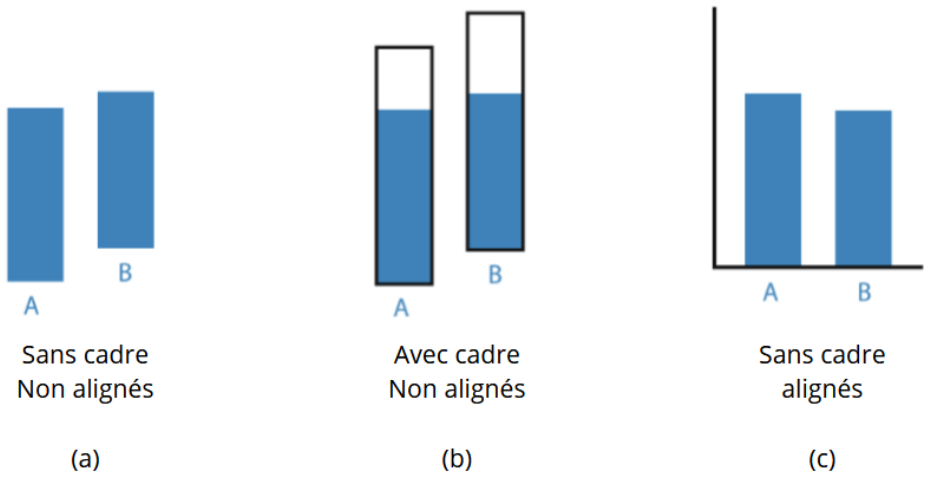


FIGURE 2.8 – Selon la loi de Weber [42], la comparaison entre longueurs repose sur des différences relatives plutôt qu'absolues. (a) Comparer les longueurs de rectangles non encadrés et non alignés de tailles légèrement différentes s'avère difficile. (b) L'ajout d'un cadre permet de comparer les tailles très distinctes des rectangles non remplis entre le sommet de la barre et le cadre. (c) L'alignement des barres facilite également le processus de jugement (graphiques tirées de [3] (2014, ©Taylor and Francis Group LLC)).

le facteur de mensonge, qui examine l'amplitude perçue d'une variable dans la visualisation par rapport à sa valeur réelle. Le maintien d'un rapport de 1 :1 est essentiel pour éviter les distorsions, en particulier lors de l'utilisation d'échelles non linéaires ou de techniques d'agrandissement. Pour tenir compte de ces facteurs, il est recommandé de réduire la « pollution », c'est-à-dire les éléments décoratifs inutiles dans les représentations visuelles. Bien que cela s'applique principalement aux diagrammes et aux infographies à but unique, il convient d'être prudent lors de l'inclusion d'artefacts visuels supplémentaires, même s'ils offrent des références ou une redondance. Néanmoins, ce dernier point doit être nuancé, car les marqueurs redondants pourraient potentiellement aider les utilisateurs à représenter et à comparer efficacement les tailles dans le graphique.

Le mantra de Shneiderman, « Overview First, Zoom and Filter, Details on Demand » [43], est fréquemment cité dans la littérature de visualisation. Il met en évidence l'équilibre crucial entre l'exigence d'une vue d'ensemble des données et la nécessité d'explorer des détails spécifiques, la réduction des données et la navigation jouant un rôle essentiel dans la prise en charge de ces deux aspects. Toutefois, ce mantra peut s'avérer plus efficace avec des ensembles de données de taille modérée. Lorsqu'il s'agit d'ensembles de données

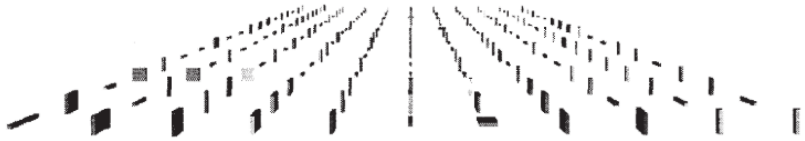


FIGURE 2.9 – Influence de la perception 3D sur la visualisation des données [3] (2014, ©Taylor and Francis Group LLC). Elle peut influencer la façon dont l'utilisateur perçoit les rapports entre distances, ce qui peut fausser les visualisations.

plus grands, comme ceux en cybersécurité, il n'est pas toujours pratique de créer une vue d'ensemble complète pour l'exploration descendante. Une autre solution pourrait alors consister à adopter le mantra de Van Ham et Perer [44] « Search, Show Context, Expand on Demand », où les résultats d'une requête servent de point de départ à la navigation.

En se basant sur les travaux de Bertin [33], Cleveland et McGill [42], Mackinlay [45] et MacEachren [46] ont examiné en détail les diverses variables disponibles dans les représentations visuelles et ont comparé leurs capacités à transmettre des informations efficacement. La figure 2.10 donne des exemples de ces variables. Chaque variable visuelle possède des capacités d'expression différentes et fonctionne plus efficacement avec des types de données spécifiques. La sélection des variables visuelles joue un rôle essentiel dans la détermination de l'efficacité d'une représentation visuelle. Des chercheurs comme Mackinlay [45] et, plus récemment, Heer et Bostock [47] ont proposé des correspondances appropriées entre les données et les variables visuelles. Par exemple, la taille est bien adaptée aux données numériques, tandis que la forme convient mieux aux données catégorielles.

Des techniques d'animations ou de « snapshot » sont parfois utilisées pour représenter une modification dans les données (généralement au cours du temps). Cependant, la charge cognitive associée à l'usage de la mémoire de travail permettant de se souvenir des états précédents de la visualisation et les comparer avec l'état actuel est plus importante que la charge cognitive associée à la comparaison de deux états différents visibles en même temps.

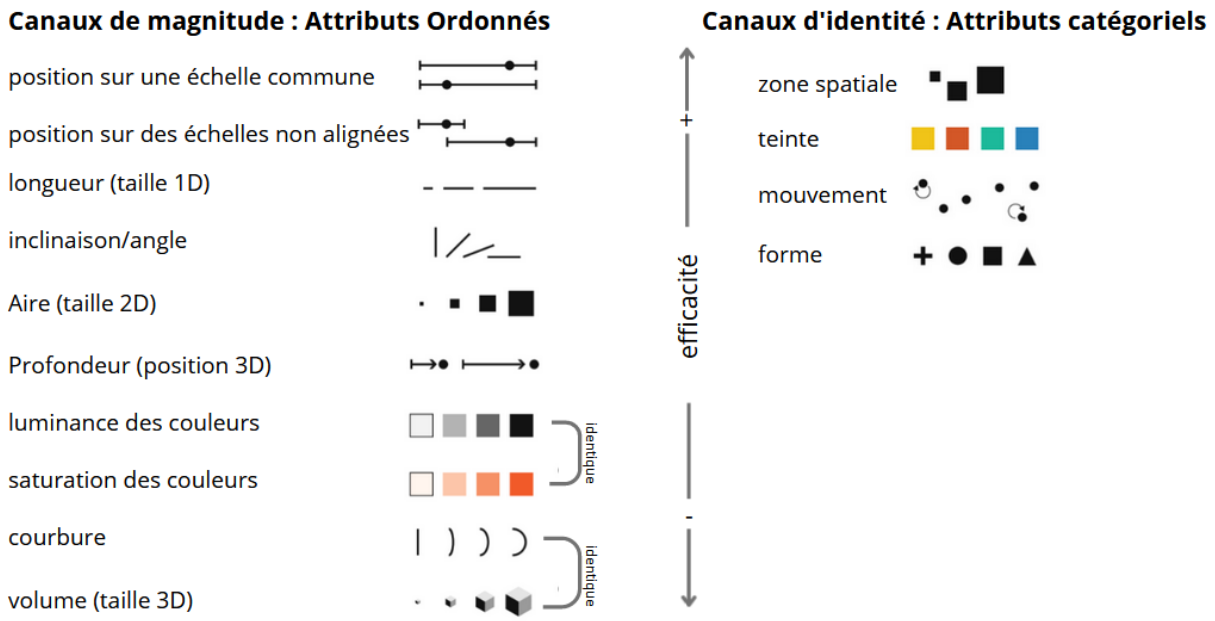


FIGURE 2.10 – Classement des canaux de visualisation selon leur efficacité en fonction des données et du type de canal [3] (2014, ©Taylor and Francis Group LLC).

2.2 Visualisations pour la cybersécurité

Nous avons vu que les visualisations de données permettent d'améliorer la compréhension de jeux de données complexes lorsqu'elles sont bien réalisées. Celles-ci sont de plus en plus intégrées dans les outils de sécurité, leur utilisation prédominante est centrée sur le rapport ou la transmission d'informations à des personnes manquant d'expertise. Étonnamment, dans le domaine des SOCs, les visualisations restent sous-utilisées malgré les avantages théoriques qu'elles offrent par rapport aux outils console et tabulaires. Dans cette partie, nous faisons un état de l'art de la recherche en visualisation spécifiquement dédié à la réponse aux alertes de cybersécurité. Nous utilisons les classifications introduites précédemment pour décrire les visualisations applicables aux environnements des SOCs. De plus, en nous appuyant sur une typologie des tâches de cybersécurité, nous établissons des liens potentiels entre ces tâches et les systèmes de visualisation correspondants. Une fois cette tâche accomplie, nous évaluons les mérites et les limites de ces visualisations en utilisant les bonnes pratiques établies précédemment. Enfin, nous nous pencherons sur les conditions techniques préalables pour des visualisations efficaces adaptées au domaine de la cybersécurité.

2.2.1 Méthodologie

Bien qu'il existe dix-huit études de littérature sur les visualisations pour la cybersécurité [48], aucune ne nous permettait de relier les visualisations cyber à d'autres domaines. En effet, les revues de littérature qui s'intéressent aux visualisations pouvant convenir à des opérateurs de SOC se concentrent surtout sur l'évaluation des visualisations et des interactions en fonction du contexte de cybersécurité et des types de données (logs, etc.). Cependant, ces revues de littérature ne détaillent pas les systèmes en instances de visualisation et ne prennent pas en compte les tâches des utilisateurs lorsqu'ils les décrivent. Cela ne permet pas de corréler une visualisation à une tâche et des données précises. En conséquence, il est impossible de justifier l'usage d'une visualisation par rapport à une autre. Cela rend l'usage de ces revues de littérature impossible pour imaginer des visualisations plus efficaces pour la cybersécurité en Immersive Analytics.

Nous nous basons sur le symposium Vizsec [1] pour construire notre corpus. En effet, celui-ci est considéré comme la conférence centrale pour les visualisations de cybersécurité [49]. Nous complétons notre corpus avec les travaux cités par Jiang et al. [48], puis nous utilisons une méthode « boule de neige » pour identifier le reste des articles pertinents. Nous identifions un total de 233 articles de visualisations pour la cybersécurité. Pour catégoriser les tâches de sécurité auxquelles correspond chaque article, nous utilisons les catégories fournies par Shiravi et al. [50] et Komadina et al. [49]. Pour sélectionner les articles pertinents, nous utilisons d'abord le titre et le résumé. Puis, nous filtrons les articles qui ne fournissent pas de cas d'utilisation et d'évaluation, ou dont la description de ces cas d'usages n'est pas suffisamment claire pour comprendre le processus de visualisation des utilisateurs (basé sur la méthode de sélection de [49]). Enfin, nous sélectionnons seulement les articles de visualisations qui permettent de répondre aux alertes, et non juste d'observer l'état d'un système, puisque c'est la base du travail des opérateurs SOCs. Finalement, nous avons retenu 50 articles (tableaux en annexes).

Pour effectuer cette analyse, nous classifions d'abord chaque article en fonction du contexte d'utilisation cyber en adoptant les typologies décrites précédemment. Ensuite, nous nous appuyons sur les descriptions des jeux de données pour définir le « what ». Les visualisations et les interactions présentées dans l'article, ainsi que les figures, permettent de décrire le « how ». Nous catégorisons également la manière dont la baseline est représentée, car la déviation par rapport au comportement nominal est très importante, ainsi que la coordination des différentes vues entre elles. Pour le « why », nous nous appuyons

1. https ://vizsec.org/

sur la description du système et sur les cas d'utilisation décrits dans l'article.

Tous les systèmes de visualisation étudiés utilisent des jeux de données de cybersécurité complexes, ce qui pousse les utilisateurs à réaliser plusieurs tâches pour enquêter sur des alertes. Ainsi, chaque système utilise plusieurs instances de visualisation qui ont pour but de représenter différentes parties du jeu de données afin d'aider l'utilisateur à accomplir différentes tâches correspondant à différentes étapes du processus d'investigation. Pour décrire efficacement les systèmes en utilisant le cadre de Munzner, nous nous sommes appuyés sur le mantra de Van Ham et Perer [44] et le processus de traitement des alertes décrit dans le chapitre 1 pour définir cinq étapes que les systèmes de visualisations pour la cybersécurité permettent de suivre : sélection des anomalies, compréhension des anomalies, développement à la demande, recherche de propagations, prise d'action. Pour chaque étape, nous déterminons le what/why/how et les interactions qui permettent de passer à l'étape suivante. De plus, nous catégorisons aussi les interactions avec les systèmes de visualisations et les méthodes utilisées pour coordonner les différentes visualisations entre elles.

2.2.2 Caractérisation des tâches dans le domaine de la cybersécurité des SOCs

Il existe différentes catégories de tâches de cybersécurité que peuvent effectuer les opérateurs des SOCs. Dans cette section, nous détaillons les valeurs que ces catégories peuvent prendre. Il arrive souvent qu'un système de visualisation corresponde à plusieurs catégories en même temps. Il peut par exemple servir à surveiller les menaces d'initiés (menaces venant du personnel interne) ET les activités du réseau en même temps. Pour les catégories de domaine (c'est-à-dire les tâches de cybersécurité), nous avons adapté la classification proposée par Komadina et al. [49] : une visualisation de cybersécurité peut aider à comprendre les règles (algorithmes de détection, pare-feu), surveiller les menaces d'initiés, surveiller les vulnérabilités, surveiller les comportements de rootages, et surveiller les réseaux. Nous avons utilisé la typologie de Shiravi pour détailler davantage le type de réseau [50] visualisé : hôte/serveur, interne/externe, activité des ports, modèles d'attaque, comportement de routage.

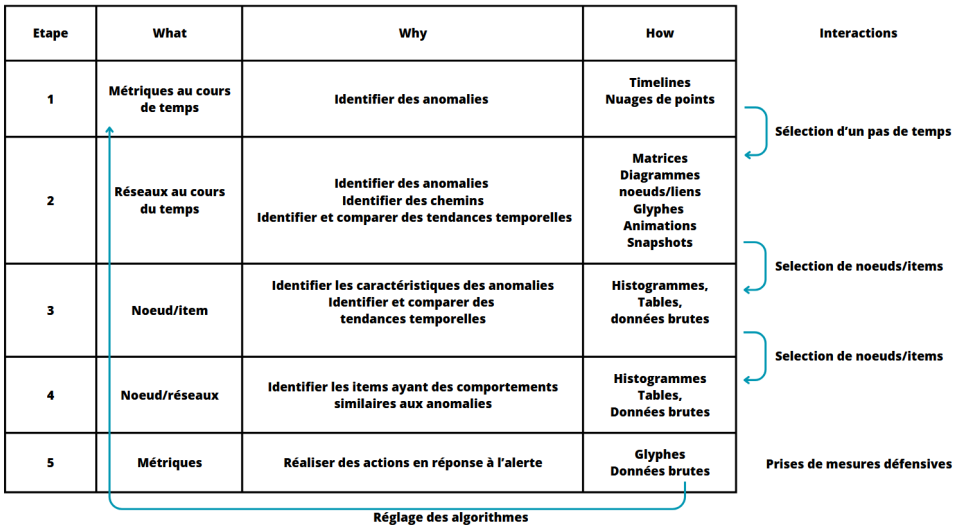


FIGURE 2.11 – Description du processus d'utilisation des visualisations pour la cybersécurité.

2.2.3 Description des systèmes de visualisations pour la réponse aux alertes de cybersécurité

Parmi les différentes catégories de tâches de cybersécurité examinées, les visualisations liées à la description des règles de pare-feu présentent un processus de travail distinct par rapport aux autres. Les visualisations des règles de pare-feu se concentrent en effet sur l'identification des chevauchements entre plusieurs règles et ont pour but de faciliter la compréhension des opérations complexes de pare-feu, qui impliquent souvent un nombre important de règles, atteignant parfois plusieurs milliers. Ces visualisations utilisent des diagrammes « sunburst », des coordonnées parallèles et des représentations basées sur le volume pour transmettre efficacement les relations entre les règles du pare-feu.

La plupart des autres visualisations du corpus (48/50) s'inscrivent dans le flux de travail commun (Figure 2.11) que nous avons défini précédemment : sélection des alertes, identification des anomalies, exploration des détails, recherche de propagation et prise d'actions (deux exemples Figure 2.12 et Figure 2.13). La majorité de ces visualisations (65 %) ont pour objectif de comprendre le schéma d'attaque. Bien que ces articles citent le mantra de Schneiderman (« Overview, zoom and filter, and detail on demand ») comme source d'inspiration, néanmoins, ce flux de travail est plus proche de celui décrit par Van Ham et Perer « Search, Show Context, Expand on Demand », où les résultats de la recherche servent de point de départ à la navigation. Bien que les systèmes de visualisations

FIGURE 2.12 – Vues principales de Situ [51] (2019, ©IEEE). L'utilisateur commence par observer les métriques au cours du temps pour identifier des anomalies à l'aide de timelines (a). Il sélectionne un pas de temps afin d'avoir accès au diagramme nœuds/liens (b) qui représentent le réseau pour rechercher les tendances de communication. Une fois qu'il identifie des nœuds suspects (ceux qui ont un comportement aberrant), il les sélectionne pour une analyse plus détaillée avec des histogrammes (c) et des données brutes (non visibles ici).

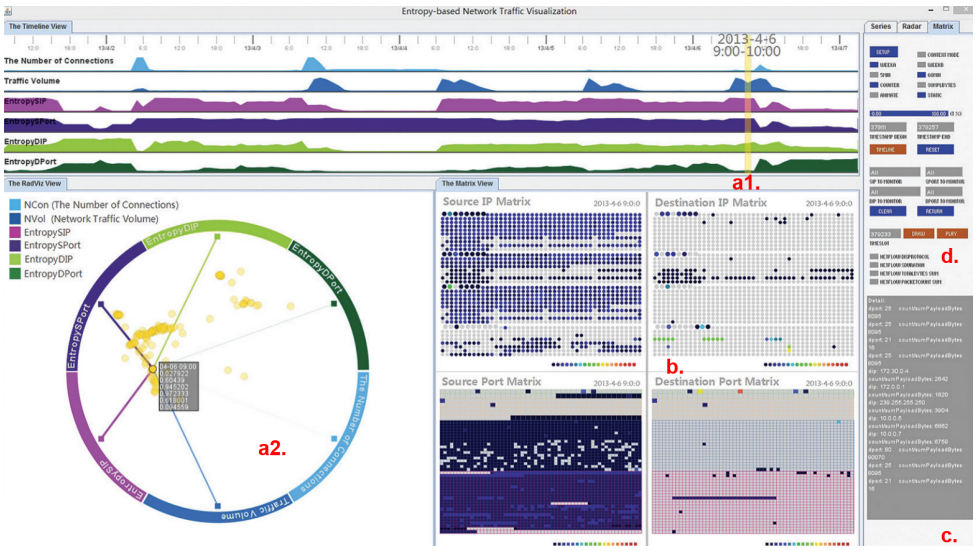


FIGURE 2.13 – Vues principales de Entvis [52] (2015, ©IEEE). L'utilisateur utilise des timelines (a1) et un nuage de points (a2) pour identifier les pas de temps suspects. Après la sélection d'un pas de temps, il utilise des matrices (b) pour identifier les communications au sein du réseau pour détecter les ports suspects (ceux qui ont un comportement aberrant). En sélectionnant un port, des informations supplémentaires sont disponibles sous forme de données brutes (c). À noter la présence des interactions abstraites en haut à droite (d).

étudiés s'inscrivent dans ce flux, ils ne proposent pas tous les cinq étapes. Ainsi, la majorité (95 %) d'entre eux contiennent des représentations de la deuxième étape, et près de la moitié (48 %) proposent des représentations des trois premières étapes. En s'inspirant du cadre de Munzer, nous pouvons décrire ces étapes comme suit :

Sélection des alertes/anomalies (72 % des articles totaux) – Dans un premier temps, les opérateurs se concentrent sur la sélection des alertes qu'ils souhaitent traiter. Trois articles utilisent des représentations basées sur des tickets pour les alertes émises par un système de détection d'intrusion (IDS). Les autres visualisent des métriques au cours du temps afin de détecter des anomalies par rapport à un comportement nominal. Ces métriques peuvent aller de mesures simples, comme le nombre de communications au fil du temps, à des indicateurs plus complexes, comme l'entropie moyenne du réseau. Pour représenter ces métriques au cours du temps, les visualisations majoritaires sont les histogrammes (53 %). Il est intéressant de noter que trois articles utilisent des représentations cycliques comme des spirales ou des cercles concentriques pour mettre en valeur l'aspect périodique des métriques des réseaux. 20 % des articles utilisent un IDS pour

générer des alertes. Une fois que les opérateurs ont identifié les anomalies qui méritent d'être étudiées, ils passent à l'étape suivante en sélectionnant l'anomalie ou l'intervalle de temps correspondant.

Compréhension des anomalies (94 % des articles totaux) – La deuxième étape consiste à représenter le réseau et les utilisateurs concernés par les alertes dans le laps de temps choisi. L'objectif de l'opérateur est de comprendre les anomalies identifiées en examinant les valeurs aberrantes et les tendances temporelles. En général, l'opérateur cherche à identifier les éléments ou les nœuds du réseau qui s'écartent de leur comportement normal, par exemple un utilisateur qui consomme davantage de données que d'habitude. Les visualisations sous forme de diagrammes nœuds/liens(35 %), de matrices (21 %), d'histogrammes (19 %) et de coordonnées parallèles (17 %) sont couramment utilisées à cette fin. Les activités des nœuds au cours du temps peuvent être représentées à l'aide de glyphes (10 %), bien que la majorité des visualisations utilisent de snapshots (20 %) ou des animations (15 %) pour le faire. Lorsque l'opérateur détecte une valeur aberrante, il sélectionne l'élément correspondant pour passer à la troisième étape.

Exploration des détails (60 % des articles totaux) – La troisième étape consiste à confirmer les soupçons de l'opérateur en examinant des caractéristiques détaillées des éléments sélectionnés, telles que le protocole de communication ou les informations d'identification de l'utilisateur, afin de déterminer si leur comportement est en effet suspect. En règle générale, l'opérateur s'appuie sur le comportement nominal de l'élément en question à des fins de comparaison. Les données brutes (50 %) et les tableaux(23 %) sont utilisés (parfois ensemble) pour représenter les caractéristiques spécifiques, tandis que des histogrammes (47 %), décrivent le comportement dans le temps pour visualiser le comportement nominal et tout écart potentiel. À ce stade, l'opérateur peut choisir de sélectionner un sous-réseau composé de nœuds adjacents pour passer à la quatrième étape.

Recherche de propagation (20 % des articles totaux) – Au cours de la quatrième étape, l'opérateur cherche à déterminer si l'attaque a réussi à atteindre et à compromettre d'autres nœuds. Pour ce faire, il représente un sous-réseau du réseau de la deuxième étape à l'aide des mêmes techniques de visualisation. L'objectif de l'opérateur est d'identifier les similitudes dans les comportements des nœuds, qui peuvent indiquer qu'ils ont été compromis.

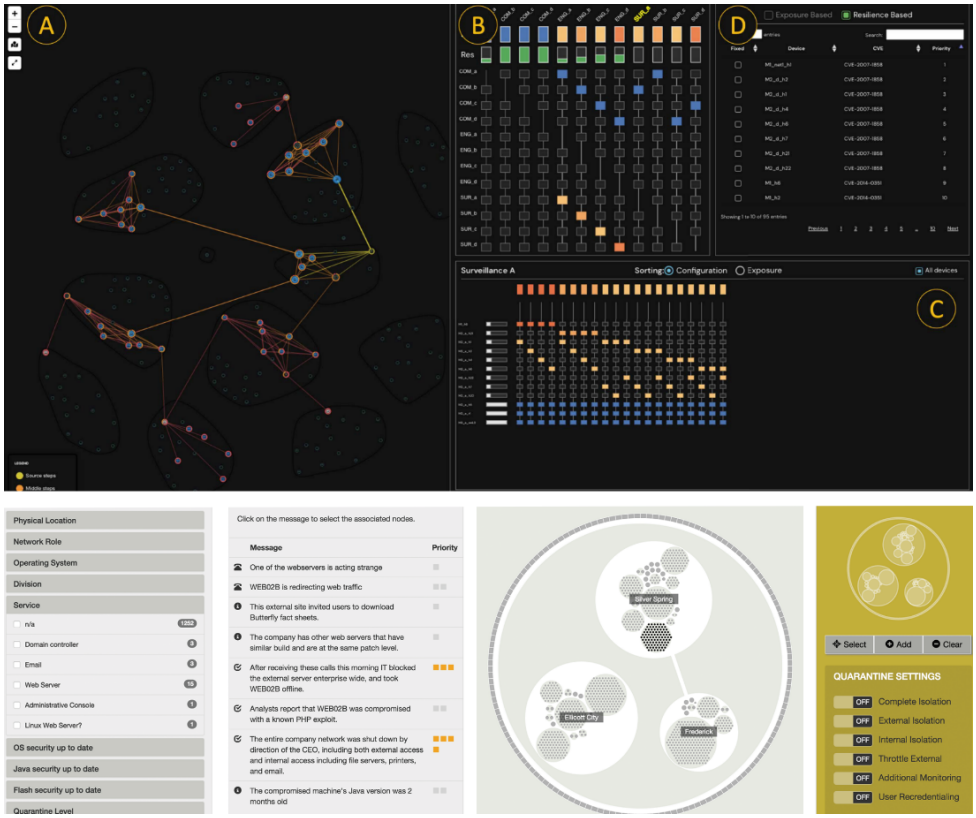
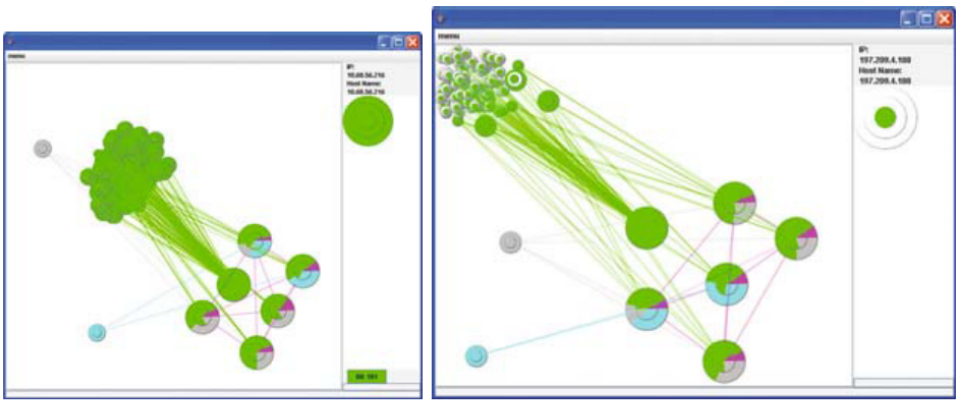


FIGURE 2.14 – Exemples d'interfaces pour la prise de mesures défensives. En haut, BUCEPHALUS [53] (2021, ©IEEE) permet de mettre en œuvre des mesures proposées par le système (D). En bas, Ocelot [54] (2015, ©IEEE) permet de mettre en quarantaines les ordinateurs sélectionnés par l'opérateur.

Prise d'actions (15 % des articles totaux) – Une fois que l'opérateur a identifié et compris l'attaque, il peut prendre les mesures appropriées pour y remédier (Figure 2.14), par exemple en arrêtant l'ordinateur compromis. Ces actions sont présentées visuellement à l'aide de glyphes superposés aux visualisations utilisées lors de l'étape deux. En outre, il est possible d'accéder à des informations plus détaillées sur les actions par le biais d'une autre vue qui affiche des données brutes. Après avoir traité l'attaque, l'opérateur peut utiliser les données collectées pour affiner les algorithmes de détection. Ces ajustements d'algorithmes sont représentés dans une vue distincte consacrée à l'ajustement des algorithmes.



(a) (b)

FIGURE 2.15 – Earlman et Rheingans [55] (2007, ©Springer) utilisent des glyphes pour représenter le comportement temporel des nœuds.

2.2.4 Spécificités

Aspect temporel des données

Dans le contexte des étapes 1 à 4, le temps joue un rôle crucial dans les données représentées. L'écart par rapport au comportement nominal sert de critère principal pour déterminer les véritables alertes positives. Cependant, les représentations en 2D présentent des limites en termes d'espace et de quantité et de type de données pouvant être affichées. Pour surmonter ces limites, les auteurs choisissent souvent de diviser leurs visualisations en plusieurs vues, chacune représentant l'une des étapes susmentionnées. Des timelines permettent d'afficher efficacement les mesures de métriques dans le temps pour les première et troisième étapes. En revanche, elles ne mettent pas en valeur des comportements périodiques qui pourraient indiquer des activités suspectes. Des représentations cycliques peuvent être utilisées pour mettre en valeurs des comportements périodiques, mais ont le défaut de nécessiter de grands espaces de visualisation, car elles doivent utiliser trois dimensions de l'espace de visualisation (deux pour afficher la série temporelle sous forme cyclique et la troisième pour afficher sa valeur) pour afficher une série temporelle bidimensionnelle. Il peut donc être difficile de corréler d'autres informations avec ces représentations.

De plus, la représentation des connexions temporelles d'un réseau dans un espace 2D devient un véritable défi. C'est pourquoi des techniques telles que les glyphes et les animations sont fréquemment utilisées (Figure 2.15). Toutefois, ces techniques font appel à la mémoire à court terme et peuvent être exigeantes sur le plan cognitif pour les opéra-

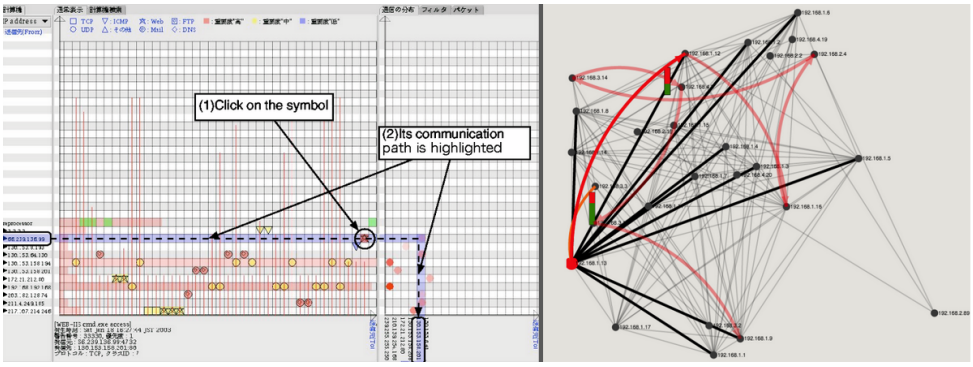


FIGURE 2.16 – Exemples de superposition. À gauche, Snortview [56] (2004, ©ACM) superpose les chemins d'accès avec les alertes sur une représentation matricielle. À droite, Percival [57] (2015, ©IEEE) superpose les chemins d'attaques et les contre-mesures disponibles sur un diagramme nœuds/liens .

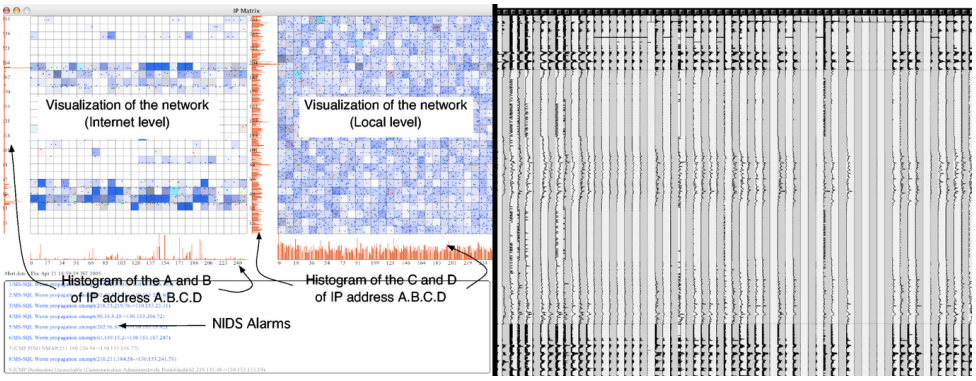


FIGURE 2.17 – Exemples de Juxtaposition. À gauche, IP Matrix [58] (2013, ©ACM) juxtapose le volume de communication des ports avec la matrice représentant les communications. À droite, Koike et al. [59] (2005, ©IEEE) juxtaposent des séries temporelles pour faciliter leur comparaison.

teurs. En outre, les glyphes occupent de l'espace dans la visualisation, ce qui rend leur comparaison difficile. Une autre limite est le passage constant d'une étape à l'autre, ce qui peut être difficile pour les opérateurs, car il faut changer de perspective et s'adapter à différentes visualisations, compte tenu de la quantité importante de données affichées.

Interactions et multivues

La majorité (80 %) des papiers étudiés proposent un système de vues multiples, en utilisant majoritairement un système multiforme (64 %) pour représenter différents aspects des données. De plus, pour assister le flux de travail décrit précédemment, la majorité

des systèmes proposent une forme d'overview + détail (60 %) qui permet de présenter le contexte général des données dans une fenêtre tout en fouillant plus précisément dans une autre. Pour coordonner ces multiples vues, les systèmes utilisent des méthodes de superpostion (17 %) et de mise en évidence (23 %), qui peuvent nécessiter d'ajouter des variables visuelles aux visualisations (Figure 2.16), ce qui les rend plus encombrées. De même, créer des liens entre les données ou utiliser des techniques de brushing nécessitent des modifications des visualisations, ce qui entraîne des pertes d'informations visuelles. Pour pallier ces limitations, la juxtaposition est une stratégie intéressante (employée par 28 % des systèmes) qui permet de relier ensemble les données grâce à leurs positions relatives sur un axe commun (Figure 2.17).

Le choix d'utiliser des vues multiples se reflète dans les systèmes d'interactions utilisés : la sélection (simple ou multiple) de points pour accéder à ces informations dans une autre fenêtre (68 %). Une autre façon d'accéder aux détails des données est d'utiliser des tooltips servant de sources d'information succinctes pour un ou plusieurs nœuds, bien que leur taille limite les informations fournies, ce qui incite à l'utilisation de fenêtres supplémentaires contenant des données brutes. Enfin, dû à la complexité des données et de leur structure, la plupart des systèmes (89 %) proposent un système abstrait pour les filtrer et effectuer des requêtes.

Espace de travail cyber et collaboration

Le concept de juxtaposition peut être étendu à un espace de travail où le positionnement stratégique permet de relier différents éléments. Bien que Singh et al. [61] et Fink et al. [23] explorent des notions similaires, ils se concentrent principalement sur le maintien de la continuité du flux de travail de l'opérateur pour la rétention de la mémoire plutôt que sur l'établissement de liens directs entre les visualisations. Si des espaces de travail comme celui de Schufrin et al. [62] proposent une augmentation du nombre de visualisations, ils tendent à les organiser dans des fenêtres plutôt qu'à fournir des moyens de les relier entre elles.

Les aspects liés à la collaboration, comme présenté dans AOH [7] et OCEAN [60], sont également pris en compte dans ces représentations, offrant aux analystes non seulement la possibilité de se remémorer leurs raisonnements, de reprendre leur travail ou de créer des rapports, mais aussi de collaborer efficacement. Par exemple, le logiciel OCEAN introduit une représentation innovante (Figure 2.18) des alertes traitées par une équipe d'opérateurs sous forme de « diagramme de problème » et intègre également des visualisations spéci-

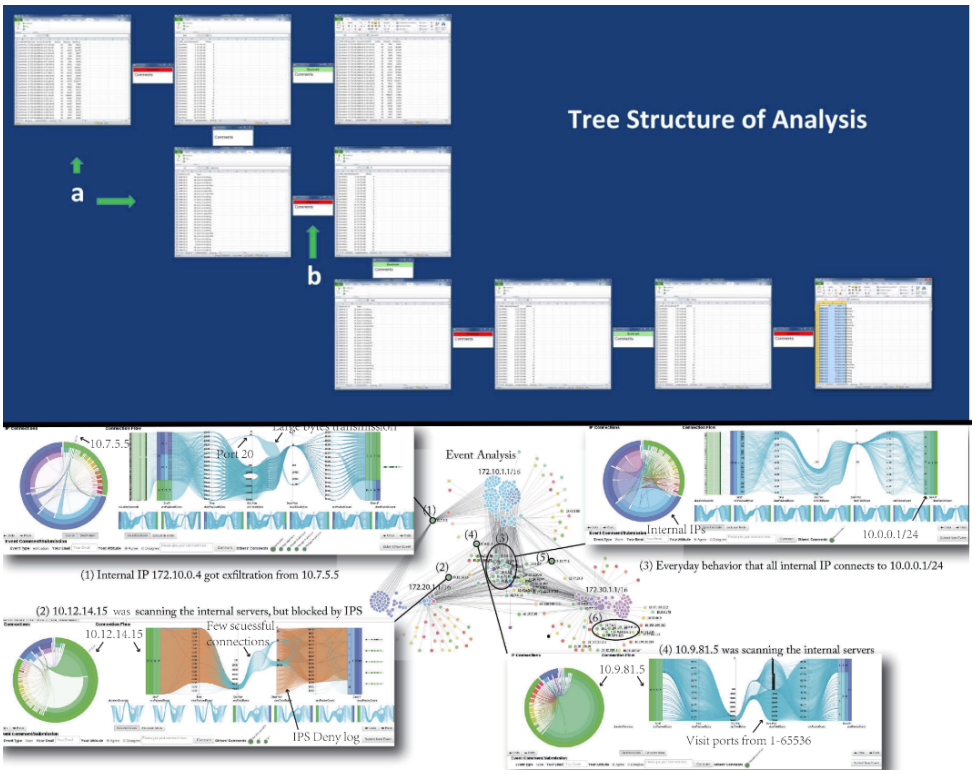


FIGURE 2.18 – En haut, l'espace de travail de Fink et al. [17] (2009, ©IEEE) permet d'organiser des tableaux en laissant les opérations pour passer de l'un à l'autre visibles, ce qui permet de retrouver le chemin mental effectué par l'opérateur. En bas, OCEAN [60] (2014, ©ACM) organise les différents inputs des opérateurs sous forme d'un diagramme nœuds/liens collaboratif.

fiques à la collaboration au sein des fenêtres principales. En ce qui concerne le contexte chronologique, plusieurs des visualisations susmentionnées utilisent un système de cartes mentales. Par exemple, des systèmes comme MAD [63] se concentrent sur la présentation d'un « snapshot » d'un réseau spécifique afin d'établir une conscience contextuelle des alertes.

2.3 Conclusion

Malgré l'efficacité de visualisations de données dans de nombreux domaines pour simplifier la compréhension de données complexes, celles-ci sont peu adoptées par les opérateurs cyber. Pour comprendre ce manque d'adoption, nous avons utilisé le cadre de Munzner pour décrire les systèmes de visualisations pour la cybersécurité, en fonction des jeux de données représentés, des tâches des opérateurs, et des techniques de visualisations et interactions utilisées. Ainsi, nous avons explicité le processus détaillé des tâches qu'effectuent les opérateurs utilisant les visualisations cyber, conformément au cadre théorique de Van Ham et Perer [44]. L'opérateur commence par l'examen des alertes ou des métriques, puis se concentre sur des intervalles de temps spécifiques ou des régions importantes, étudie les schémas inhabituels et affine le champ d'application pour repérer les cibles exactes. Après l'exécution de l'attaque, l'analyse se poursuit pour identifier les impacts, développer des contre-mesures et améliorer les algorithmes de détection. Cela montre la nécessité de dépasser le mantra de Schneidermann, qui ne convient pas aux ensembles de données cyber trop volumineux [3] comme ceux visualisés en cybersécurité. Ainsi, les visualisations cyber suivent un modèle qui ne correspond pas aux données qu'elles représentent. De plus, les systèmes utilisent majoritairement des visualisations simples et communes, organisées dans des systèmes multivues qui nécessitent de nombreuses interactions. Bien que ces visualisations soient simples à prendre en main, leur adaptation aux représentations des données en cybersécurité s'avère limitée. En effet, ces visualisations sacrifient souvent la représentation temporelle au profit d'autres données telles que les métriques ou la structure des réseaux, et recourent à des techniques exigeant des efforts de la mémoire de travail, comme les animations. Cette limitation pose un problème crucial, car la comparaison entre l'état actuel d'un réseau et son comportement nominal, fondamentale pour les opérateurs en cybersécurité, est ainsi entravée. De plus, le traitement du volume important et de la complexité des données conduit à l'utilisation de systèmes multivues qui requièrent de nombreuses variables visuelles, pouvant compliquer la visualisation et

la rendre fastidieuse. Cette complexité se traduit également par un recours fréquent à de multiples interactions pour filtrer les données et gérer ces vues, ce qui peut être fastidieux pour les opérateurs. Ainsi, les visualisations en cybersécurité ne parviennent pas à soutenir les modèles mentaux des opérateurs, élément essentiel pour des visualisations efficaces [64]. Enfin, certaines visualisations s'intéressent à l'usage de l'espace de visualisation en lui-même pour organiser les différentes vues, ce qui pourrait aider le processus de construction des sens des opérateurs ainsi que la collaboration. Mais pour l'instant ces représentations sont limitées à des mindmap 2D. Ainsi, nous identifions trois verrous des visualisations 2D pour la cybersécurité que l'Immersive Analytics pourrait aider à faire sauter. Les visualisations 2D ont des difficultés à (1) représenter des comportements périodiques au cours du temps, (2) représenter l'évolution des états d'un réseau au cours du temps, et (3) corréler efficacement plusieurs visualisations entre elles. De plus, nous identifions deux possibilités d'améliorations des solutions 2D existantes par l'Immersive Analytics : (1) l'usage de l'espace de visualisation pour la construction de sens, et (2) l'usage des systèmes de visualisation pour la collaboration.

Résumé

Dans ce chapitre, nous avons cherché à comprendre pourquoi les opérateurs cyber n'adoptent pas largement les visualisations. Pour cela, nous avons présenté le domaine de la visualisation des données, en particulier le pipeline de visualisation, qui décrit les interactions nécessaires à la transformation des données de données brutes de visualisation utilisable. Puis nous avons présenté le cadre de Munzner, qui permet de catégoriser les visualisations en fonction du type de données, des tâches effectuées, et des techniques de visualisation et d'interaction. En outre, nous avons présenté les bonnes pratiques à suivre pour créer des visualisations efficaces. En utilisant le cadre de Munzner et les bonnes pratiques précédemment citées, nous avons expliqué pourquoi les visualisations proposées par la communauté scientifique ne sont pas largement adoptées en cybersécurité. En effet, les opérations des opérateurs cyber suivent les principes du traitement des données massives de Van Ham et Perer. Ces opérations reposent principalement sur des visualisations simples et familières, intégrées dans des systèmes multi-vues nécessitant une interaction approfondie. Bien que les visualisations utilisées soient bien connues des experts en cybersécurité, elles présentent plusieurs lacunes. Nous avons observé que les visualisations 2D rencontrent des difficultés à représenter les aspects temporels et à corréler efficacement les données, deux éléments cruciaux pour la détection des signaux faibles en cybersécurité. De plus, l'utilisation de l'espace de visualisation pour l'organisation des données et la collaboration a été identifiée comme une piste prometteuse.

Chapitre 3

Immersive Analytics pour les SOCs

Dans la section précédente, nous avons décrit le processus des visualisations 2D pour la cybersécurité et identifié leurs points faibles. Certaines techniques de visualisation 3D peuvent pallier ces faiblesses. Dans ce chapitre, nous présentons donc les différentes technologies d'affichage 3D immersif. Nous nous intéressons particulièrement au domaine de l'Immersive Analytics, qui concerne l'utilisation de ces technologies pour visualiser des données. Après avoir présenté le domaine et les principaux défis auxquels il est confronté, nous présentons les différents cas d'utilisation potentiels de l'Immersive Analytics en cybersécurité, ainsi que les pistes de recherche sur les solutions techniques qu'il reste à développer pour la rendre acceptable par les opérateurs des SOCs.

3.1 Réalité Mixte

La réalité mixte (RM) est un domaine interdisciplinaire qui intègre des éléments virtuels et réels pour créer des expériences immersives et interactives. Ses origines remontent aux années 1960, lorsque les premiers développements des technologies de la Réalité Virtuelle (RV) et de la Réalité Augmentée (RA) ont jeté les bases du domaine. Le terme « Réalité Mixte » a été officiellement introduit dans les années 1990 par Paul Milgram et Fumio Kishino [65], qui ont proposé la notion de « continuum de virtualité » englobant un spectre allant de l'environnement réel à l'environnement virtuel.

La Réalité Virtuelle en particulier a franchi une étape importante avec l'introduction des écrans montés sur la tête (HMD pour Head Mounted Display) dans les années 1980 et 1990. Ces casques ont prouvé leur utilité dans divers domaines tels que l'architecture, l'ingénierie, les soins de santé et l'éducation à des fins de formation, de visualisation de la conception, de simulations et d'interventions thérapeutiques. Depuis 2013, de nouveaux casques comme l'oculus rift et l'HTC vive sont apparus sur le marché à destination des particuliers et ont contribué à démocratiser la Réalité Virtuelle (Figure 3.1).

Selon la première édition du Traité de la Réalité Virtuelle (TRV) [66], la Réalité

Virtuelle est « un domaine scientifique et technologique exploitant l'informatique et les interfaces comportementales en vue de simuler dans un monde virtuel le comportement d'entités 3D, qui sont en interaction en temps-réel et avec un ou des utilisateurs en immersion pseudo-naturelle par l'intermédiaire de canaux sensorimoteurs ». Cette définition met en avant les aspects techniques et humains de la réalité virtuelle qui englobe à la fois les technologies immersives, mais aussi l'aspect cognitif des utilisateurs. Ce dernier doit donc être pris en compte pour créer un sentiment de présence chez l'utilisateur, c'est-à-dire qu'il sente que ses actions et réactions sont semblables à celles qu'elles auraient été dans une situation réelle.

FIGURE 3.1 – Chronologie des technologies de Réalité Mixte

Le développement d'environnements de Réalité Virtuelle (RV) a été grandement facilité par les avancées technologiques, notamment en matière de puissance de calcul, de traitement graphique et de technologies d'interaction. Par exemple, les moteurs de jeu comme Unity [1] et Unreal Engine [2] ont développé des librairies dédiées à la RV dès 2013. Ces librairies simplifient la création d'expériences immersives en fournissant des outils pour le rendu graphique, les simulations physiques et les interactions avec l'utilisateur. Ces moteurs ont réduit la complexité technique du développement des applications de Réalité Virtuelle, permettant aux créateurs de se concentrer sur la conception du contenu et des interactions dans le monde virtuel. En outre, les capacités multiplateformes des moteurs de jeu ont rendu ces applications accessibles sur différents appareils, contribuant ainsi à l'adoption plus large des expériences de RV dans divers domaines au-delà du jeu.

3.2 Immersive Analytics

L'avancée des technologies de réalité virtuelle, en particulier celle des casques, a permis d'augmenter leurs performances d'affichage, comme la résolution et le champ de vision, ce

1. https ://unity.com/fr, 02/10/2023
2. https ://www.unrealengine.com/fr, 02/10/2023

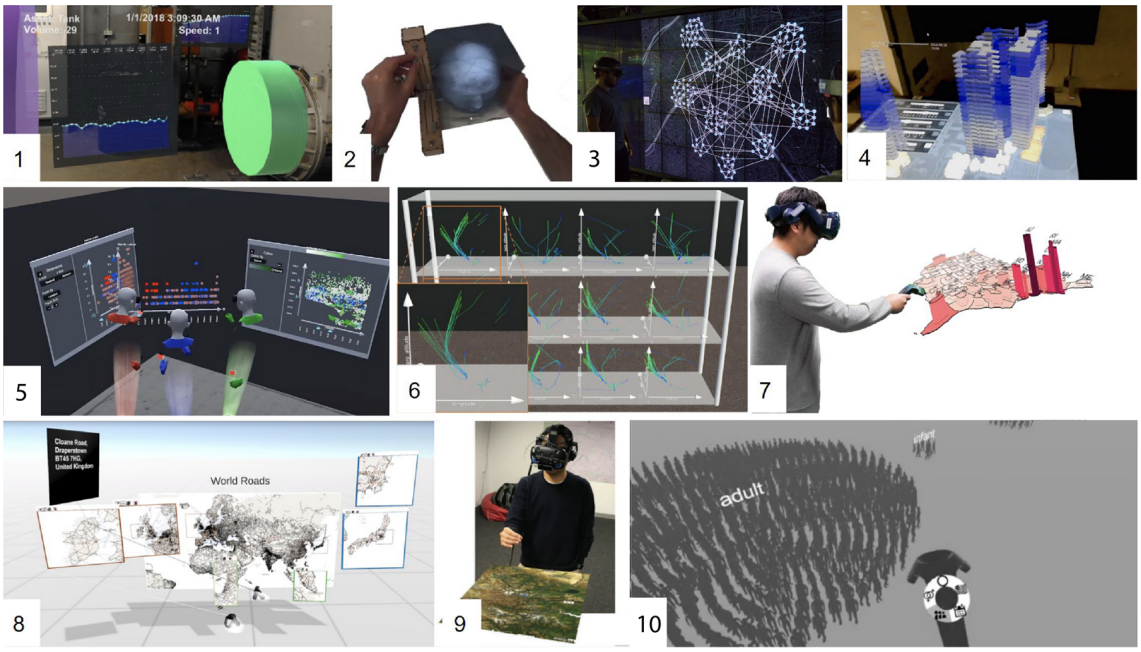


FIGURE 3.2 – Des exemples d'usages d'Immersive Analytics collectés par Ens et al.[67] (2021, ©ACM). 1) Corsican Twin [68] (2020, ©ACM) (2) Contrôleurs tangibles d'axes en RA [69] (2020, ©ACM) (3) navigation personnalisée des réseaux en RA partagée [70] (2020, ©HAL), (4) UpLift [71] (2021, ©IEEE) combine la RA, une table tactile et des objets tangibles pour l'analyse énergétique des bâtiments (5) FIESTA [72] (2019, ©ACM) permet l'analyse immersive collaborative en réalité virtuelle, (6) une étude des small-multiples en réalité virtuelle [73] (2020, ©IEEE), (7) Tilt Map [74] (2021, ©IEEE) permet de combiner des cartes géographiques avec des histogrammes (8) un système de navigation à plusieurs échelles pour la navigation sur carte en réalité augmentée [75] (2020, ©ACM), (9) des interactions à main nue permettent de naviguer dans des cartes en RA [76] (2019, ©IEEE), et (10) visualisation immersive anthropomorphique [77] (2018, ©ACM).

qui permet d'afficher une plus grande quantité de données aux utilisateurs (Figure 3.2). Les chercheurs en réalité virtuelle ont donc eu l'idée naturelle d'utiliser ces technologies dans les années 90 pour créer des systèmes de visualisation scientifique, par exemple en utilisant des systèmes 3D stéréoscopique pour visualiser des données sur des diagrammes nœuds/liens 3D [78]. Depuis les années 2000, l'essor de nouvelles solutions immersives de plus en plus performantes a suscité un intérêt accru pour l'Immersive Analytics. Cela a conduit les chercheurs à organiser le champ de recherche et à proposer plusieurs définitions :

— Chandler et Al. [2] définissent l'Immersive Analytics comme « l'étude des nouvelles

technologies d'interaction et d'affichage pouvant être utilisées pour soutenir le raisonnement analytique et la prise de décision ». Ainsi, l'objectif est de fournir des interfaces multisensorielles pour les approches analytiques qui favorisent la collaboration et permettent aux utilisateurs de s'immerger dans leurs données. Pour eux, l'Immersive Analytics « s'appuie sur des technologies telles que les grandes surfaces tactiles, les environnements de Réalité Virtuelle et Augmentée immersives, les affichages haptiques et audio, ainsi que les techniques de fabrication modernes ».

— Dans leur survey, Ens et al. [67] définissent l'Immersive Analytics comme « l'utilisation d'outils d'analyse engageants et incarnés pour soutenir la compréhension des données et la prise de décision ».

— Pour Hackarthon et Margolis [79] l'objectif de l'Immersive Analytics est de faciliter le soutien collaboratif à la prise de décision pour la gestion de systèmes complexes. L'Immersive Analytics exploite à la fois les outils d'analyse et les architectures actuelles et les combinent avec des espaces de données immersifs.

— Skarbez et al. [80] déclarent dans leur Research Agenda que l'Immersive Analytics vise à faciliter la prise de décision collaborative pour la gestion de systèmes complexes, en utilisant les outils d'analyse et les architectures actuelles, et en intégrant le raisonnement analytique aux espaces de données immersifs.

Ainsi, l'Immersive Analytics est un champ de recherche qui a pour but d'améliorer la compréhension des données et les processus de prise de décision. Pour cela, il explore l'usage des technologies d'interaction et d'affichage de pointe pour soutenir le raisonnement analytique. L'Immersive Analytics repose sur un large éventail de technologies, allant des grandes surfaces tactiles aux environnements immersifs de Réalité Virtuelle et Augmentée, qui transportent les utilisateurs dans leurs espaces de donnée et leur permettent une interaction intuitive avec les données.

3.2.1 Challenges

Comme l'Immersive Analytics est un domaine de recherche jeune, il n'a pas encore fait l'objet de méthodes de recherche unifiées. Les chercheurs se sont surtout attachés à justifier la valeur de leurs systèmes par rapport à des systèmes non immersifs classiques, plutôt que de créer un langage commun dans le but de trouver les méthodes de visualisations et d'interactions immersives optimales [81]. Depuis 2015, un livre [82], un survey [81], un Research Agenda [80] et un article [67] ont contribué à structurer le domaine

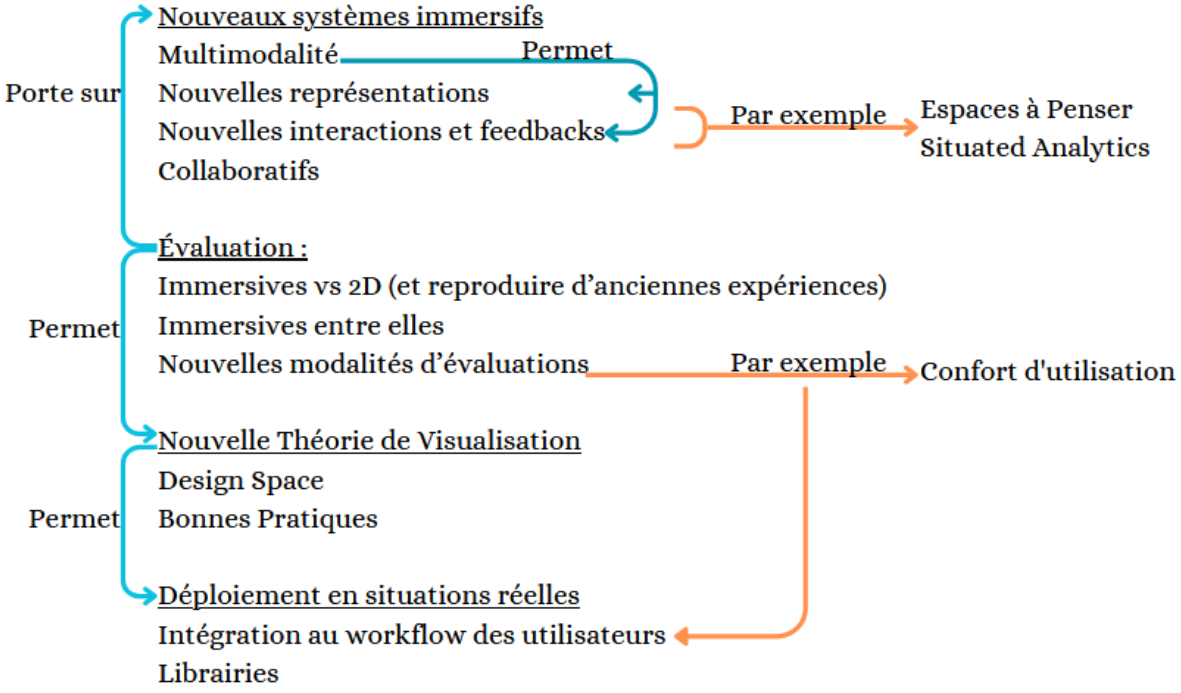


FIGURE 3.3 – Les challenges de l'Immersive Analytics

en identifiant les challenges propres aux domaines de l'Immersive Analytics. En étudiant ces travaux, nous dégageons quatre challenges pour l'Immersive Analytics qui impactent la création de nouvelles représentations de données immersives pour la cybersécurité : la création de nouvelles représentations de données immersives, l'évaluation de ces nouvelles représentations, la création d'un espace de conception et de bonnes pratiques de l'Immersive Analytics, et le déploiement en situations réelles de l'Immersive Analytics.

La recherche en Réalité Virtuelle propose de nombreuses solutions immersives qui cherchent à utiliser tous les sens de l'être humain pour immerger les utilisateurs [82]. Adopter ces systèmes multimodaux pourrait améliorer à la fois les visualisations des données et les interactions avec celles-ci. En effet, les environnements immersifs offrent un potentiel inégalé pour une immersion totale de l'utilisateur, qui peut être exploité pour représenter des données qui tirent parti de l'ensemble de l'immersion corporelle. Cette adaptation demande de trouver de nouvelles façons de représenter les données et d'interagir avec les systèmes immersifs. Alors que les représentations de données prédominantes mettent l'accent sur le canal visuel, il existe un manque notable de propositions qui intègrent le son et le retour haptique [80]. De plus, de nombreux concepts de représentation de données en 3D découlent de leurs homologues 2D, il peut être intéressant de chercher

des représentations 3D Ad Hoc [81]. Pour l'instant, les paradigmes d'interaction dans l'Immersive Analytics se basent essentiellement sur l'utilisation de contrôleurs à six degrés de liberté [83], couramment utilisés pour la sélection par rayon et la manipulation d'outils. Cependant, le domaine de l'Immersive Analytics ne se limite pas à ces interactions ou à ces contrôleurs [83]. Une façon d'interagir novatrice est d'utiliser le sentiment de présence de l'utilisateur dans l'environnement virtuel pour représenter son cheminement mental [84]. Cela pourrait permettre à des algorithmes d'apprentissage automatique de créer des relations en fonction du travail de l'utilisateur, et apporterait une nouvelle dimension au domaine de l'humain dans la boucle [67]. Une autre piste prometteuse semble être la Situated Analytics qui consiste à utiliser des représentations de données organisées en relation avec des objets, des lieux et des personnes pertinents dans le monde physique dans le but de comprendre, de donner un sens et de prendre des décisions [84]. Elle améliore la compréhension spatiale en reliant les représentations de données à des référents significatifs. Elle redéfinit la relation à l'environnement physique dans la boucle de sensemaking et permet à l'utilisateur une plus grande immersion dans son environnement et les données associées [67]. Les solutions immersives permettent également d'imaginer de nouvelles façons de collaborer dans des environnements virtuels partagés [81]. Toutefois, la relative nouveauté des technologies immersives pour une grande partie des utilisateurs peut être un frein à leur adoption [67]. La formation des utilisateurs à une collaboration efficace à l'aide de l'analyse immersive devient primordiale et comporte deux dimensions : premièrement, transmettre une compréhension approfondie de l'exécution des tâches dans le contexte immersif, y compris le partage des vues, des données et des manipulations d'objets virtuels ; et deuxièmement, cultiver une conscience des diverses voies de communication, englobant les gestes de la main, les mouvements de la tête et les signaux audio [67]. En outre, les défis s'étendent à l'adoption d'une collaboration multiplateforme, où les systèmes d'analyse immersive se heurtent aux normes de collaboration établies, allant des réunions en face à face aux visios conférences [81].

Il est important de vérifier que les nouvelles techniques d'affichages et d'interactions proposées soient évaluées en fonction des objectifs des utilisateurs, afin d'être capable de converger vers les bonnes pratiques de l'Immersive Analytics [80]. Pour cela, il faut penser une nouvelle théorie de visualisation qui modifie la façon dont est évaluée l'efficacité des visualisations, ce qui entraîne de nouvelles méthodologies d'évaluation, pour conduire vers un nouvel espace de conception adapté à l'Immersive Analytics. Élaborer une théorie globale pour l'immersion et l'engagement dans l'analyse de données et la prise de décision

dans les environnements de réalité mixte nécessite une adaptation et une extension des théories existantes, en liant l'immersion psychologique, les états émotionnels et la performance de l'utilisateur [84]. Répondre à la question de l'efficacité demande de comparer les solutions immersives entre elles, mais aussi de les comparer à des outils d'analyse visuelle plus classique [84]. Un frein à cette comparaison entre solutions immersives et non immersives est la mauvaise réputation des techniques de visualisations 3D dans la communauté de la visualisation [84]. Par exemple, les nuages de points 3D ont été jugés inefficaces [85] car plus lents qu'une représentation des nuages de points 2D, mais de nouvelles études ont montré qu'ils peuvent être plus précis [86]. En effet, des comparaisons faites il y a plus de 20 ans sur des systèmes matériels et des logiciels moins performants que les actuels avaient montré que les problèmes, comme l'occlusion, l'emportent sur les avantages apportés par la 3D [85]. Cependant, les nouvelles solutions technologiques peuvent aujourd'hui apporter des solutions à ces problèmes et de nouvelles évaluations semblent nécessaires pour réévaluer les avantages et inconvénients de l'Immersive Analytics [87]. De plus, l'évaluation des systèmes de visualisations pourrait utiliser de nouvelles métriques pour dépasser les mesures traditionnelles de la performance des tâches (le temps de complétion et le taux de succès), pour aussi prendre en compte l'engagement émotionnel et la mémorisation de l'utilisateur, qui sont renforcés par le sentiment de présence [84].

Ces évaluations contribuent à la mise en place des bonnes pratiques de l'Immersive Analytics et à la conception d'un espace de conception des techniques de représentation et d'interactions immersives efficaces [67].

Une fois que les bonnes pratiques de l'Immersive Analytics sont connues, il reste encore à faire adopter ces solutions par les utilisateurs. Pour cela, l'Immersive Analytics doit être capable de proposer des solutions qui sont à la fois faciles à créer pour les développeurs et faciles à adopter pour les utilisateurs. Les solutions d'Immersive Analytics doivent être capables de s'intégrer de façon harmonieuse dans le workflow déjà existant des utilisateurs [80]. Il est donc important pour les chercheurs en Immersive Analytics de mettre l'accent sur les scénarios réels dans leurs évaluations [80]. Ainsi, la capacité d'un système immersif à s'intégrer à un workflow préexistant peut être une nouvelle modalité d'évaluation. La conception de systèmes d'Immersive Analytics nécessite également de prendre en compte le confort des utilisateurs, surtout durant de longues sessions durant lesquelles utiliser des solutions immersives peut être plus fatigante que d'utiliser des solutions classiques [80]. Pour créer ces visualisations immersives, le développement de plateformes génériques adaptées à l'Immersive Analytics est nécessaire. Pour l'instant, la

majorité des frameworks d'Immersive Analytics reposent sur le moteur de jeu Unity [80], ce qui permet un développement rapide, mais peut entraver la propagation de ces solutions. Enfin, la démocratisation des solutions immersives entraîne des questionnements sur les risques liés aux décisions motivées par l'engagement émotionnel ainsi que sur la santé des utilisateurs de l'Immersive Analytics [84].

Synthèse

Il y a donc quatre challenges dans la recherche en Immersive Analytics qui ont un impact sur la création de nouvelles représentations de données immersives pour la cybersécurité : développer de nouvelles formes immersives de représentation des données, évaluer ces nouvelles représentations, établir un environnement de conception et des bonnes pratiques pour l'Immersive Analytics, et enfin, intégrer l'Immersive Analytics dans le monde réel. Ceux-ci sont bien sûr dépendants des limitations techniques des solutions de réalité mixte utilisées pour déployer et évaluer les systèmes de visualisations. Dans les prochaines sections, nous verrons comment l'Immersive Analytics peut améliorer les visualisations pour la cybersécurité, puis nous présenterons comment ces pistes d'améliorations s'inscrivent dans les challenges décrits plus haut.

3.3 Apport de l'Immersive Analytics pour la cybersécurité

Dans cette partie, nous présentons comment l'Immersive Analytics peut se mettre au service de la cybersécurité des SOCs. Pour cela, nous nous basons sur les limites des visualisations 2D décrites dans le chapitre 2, et nous déterminons les solutions que l'Immersive Analytics pourrait apporter pour les dépasser. Puis, nous mettons en avant les limites actuelles de l'Immersive Analytics qui empêchent son acceptation dans les SOCs. Enfin, nous concluons en faisant le lien entre les challenges actuels des visualisations pour la cybersécurité avec ceux de l'Immersive Analytics, pour identifier des pistes de recherches qui profitent aux deux domaines.

3.3.1 Constitution du corpus

Le travail effectué au chapitre 2 nous permet d'identifier les limites actuelles des visualisations pour la cybersécurité. Nous utilisons la catégorisation des tâches et des données

créée précédemment pour identifier les solutions immersives qui améliorent ces aspects. Dans un premier temps, nous cherchons les visualisations immersives qui existent dans la communauté de visualisations pour la cybersécurité. Comme ce nombre de visualisations est restreint, nous avons utilisé les travaux de Fonnet et Prié [81], Korkut et al. [88] et Besançon et al. [89] pour explorer les travaux où des solutions d'Immersive Analytics se sont montrées plus efficaces que des représentations 2D pour des problèmes similaires, c'est-à-dire avec des types de données et de tâches similaires. De plus, nous analysons les visualisations 3D non immersives proposées en cybersécurité qui ne sont pas appliquées à la réponse à alertes, mais qui pourraient s'appliquer au fonctionnement des SOCs.

3.3.2 L'Immersive Analytics pour les SOCs

Il y a quatre articles dans notre corpus qui correspondent à nos critères et qui utilisent l'Immersive Analytics. Ils se concentrent sur l'amélioration de la conscience de situation cyber partagée (shared CSA) pour les employés des SOCs, et cherchent à améliorer leur collaboration.

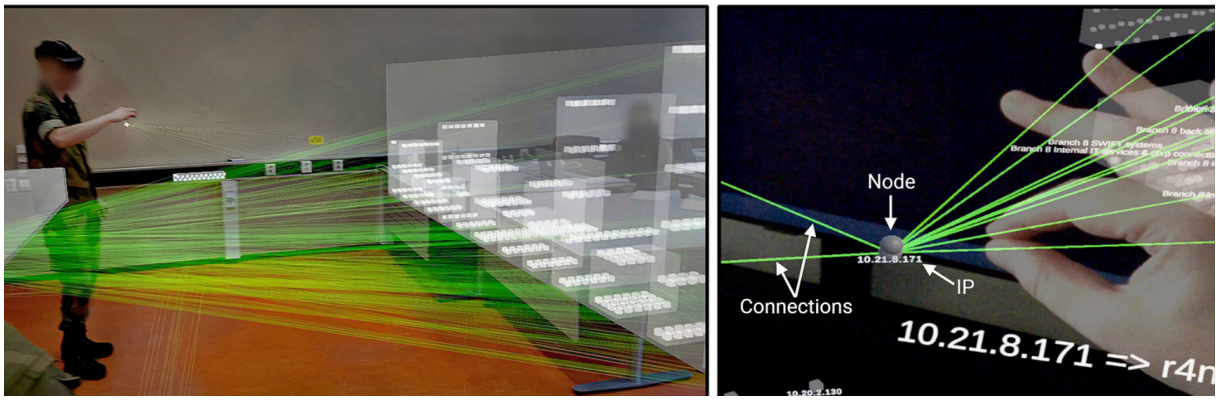


FIGURE 3.4 – Le Virtual Data Explorer (VDE) représente la topologie du réseau dans un environnement collaboratif en RA [90] (2023, ©Frontiers).

Un laboratoire de l'armée américaine a mis au point trois prototypes [91]. Le premier est le Virtual Reality Data Analysis Environment (VRDAE) qui offre un environnement collaboratif avec des outils de visualisation en 3D pour analyser les représentations de réseaux informatiques, en mettant l'accent sur l'interaction avec l'utilisateur, l'expérience et la conception de l'interface. Le second est le système visuel de détection des intrusions (VIDS) qui vise à combler le fossé entre les visualisations existantes des réseaux et de la sécurité et les futures visualisations 3D en concevant et en testant des dispositions

logiques des caractéristiques des réseaux informatiques à l'aide de la visualisation 3D. Le troisième système est l'environnement de détection virtuel (VDE) qui visualise la topologie des réseaux informatiques (Figure 3.4) et de leurs membres en combinant des formes de données 3D personnalisées pour représenter les échanges du réseau. Les tests effectués auprès d'experts en cybersécurité et d'étudiants ont montré que les visualisations 3D perceptibles par stéréoscopie sont efficaces pour comprendre le comportement nominal et la topologie des grands réseaux, mais n'ont pas été comparés avec des visualisations 2D. En outre, une étude comparant la réalité mixte 3D et les visualisations 2D a montré une amélioration de la connaissance de la situation cyber et de la communication au sein de l'équipe sans problème de performance [90].

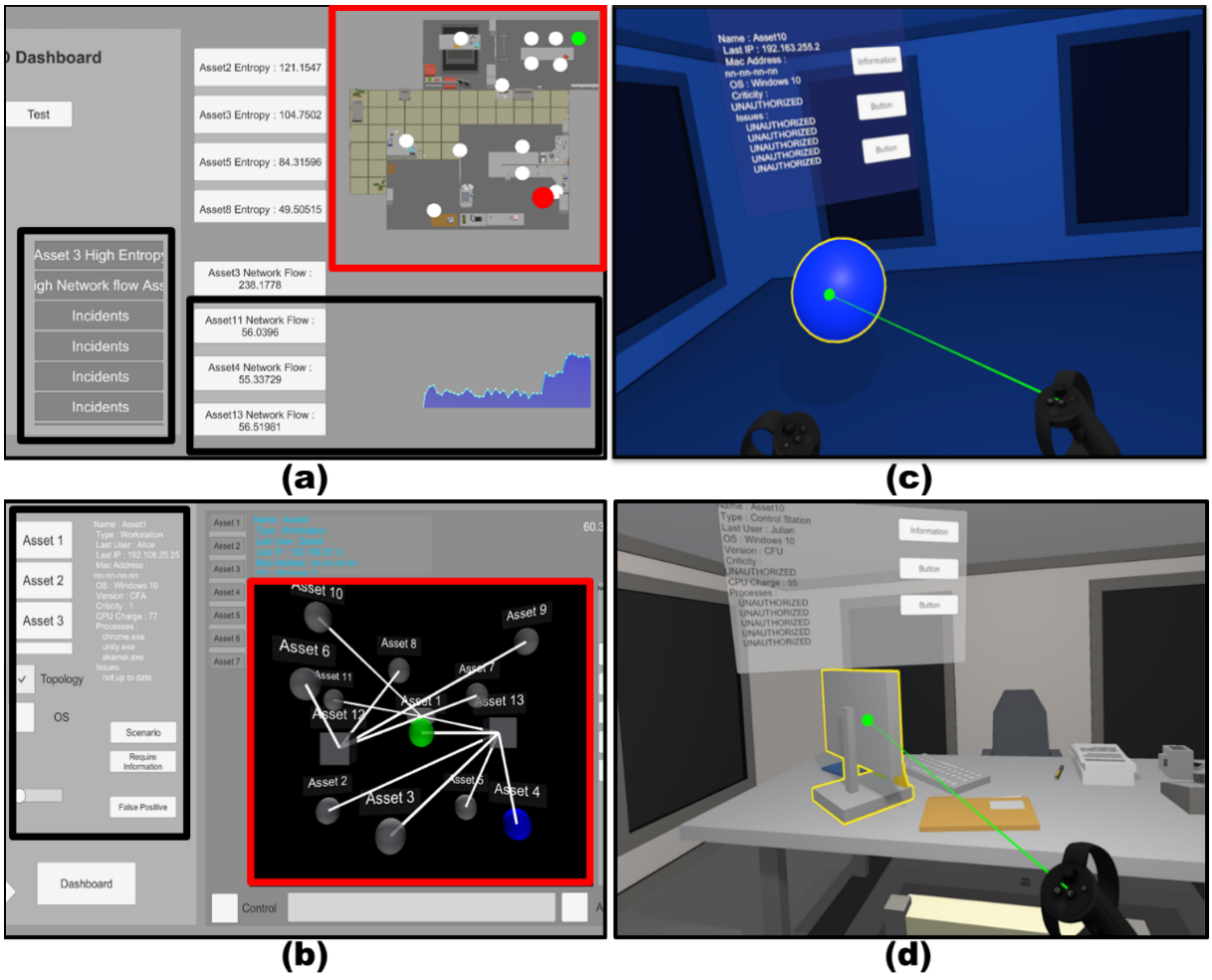


FIGURE 3.5 – La plateforme 3D CyberCOP [92] (2028, ©Springer) permet d'utiliser plusieurs vues 2D et 3D au sein d'un environement immersif pour répondre à des alertes.

Dans le même ordre d'idées, Kabil et al. ont plaidé en faveur d'une solution immersive au sein des SOCs afin d'améliorer la CSA et la collaboration suite à l'observation du manque de collaboration médiatisée entre les utilisateurs lors de visites dans des SOC français. En réponse, ils ont développé un modèle [8] et mis en œuvre un prototype : le 3D CyberCOP [92], qui intègre des visualisations de données dans un environnement virtuel basé sur l'environnement physique du réseau observé (Figure 3.5). L'objectif de la plateforme est d'améliorer la CSA des utilisateurs en fusionnant des visualisations de données immersives avec des approches de jeux sérieux tout en adaptant les pratiques collaboratives existantes en matière de cybersécurité. Un scénario d'analyse d'alertes basé sur le ransomware WannaCry est présenté comme un cas d'utilisation. Une étude d'utilisabilité impliquant 30 utilisateurs non experts a été menée avec trois environnements virtuels : abstrait, réel et hybride (une combinaison des deux autres). Les trois environnements ont été jugés utilisables, et les utilisateurs ont exprimé une préférence pour l'environnement hybride, car il était le plus pratique selon eux.

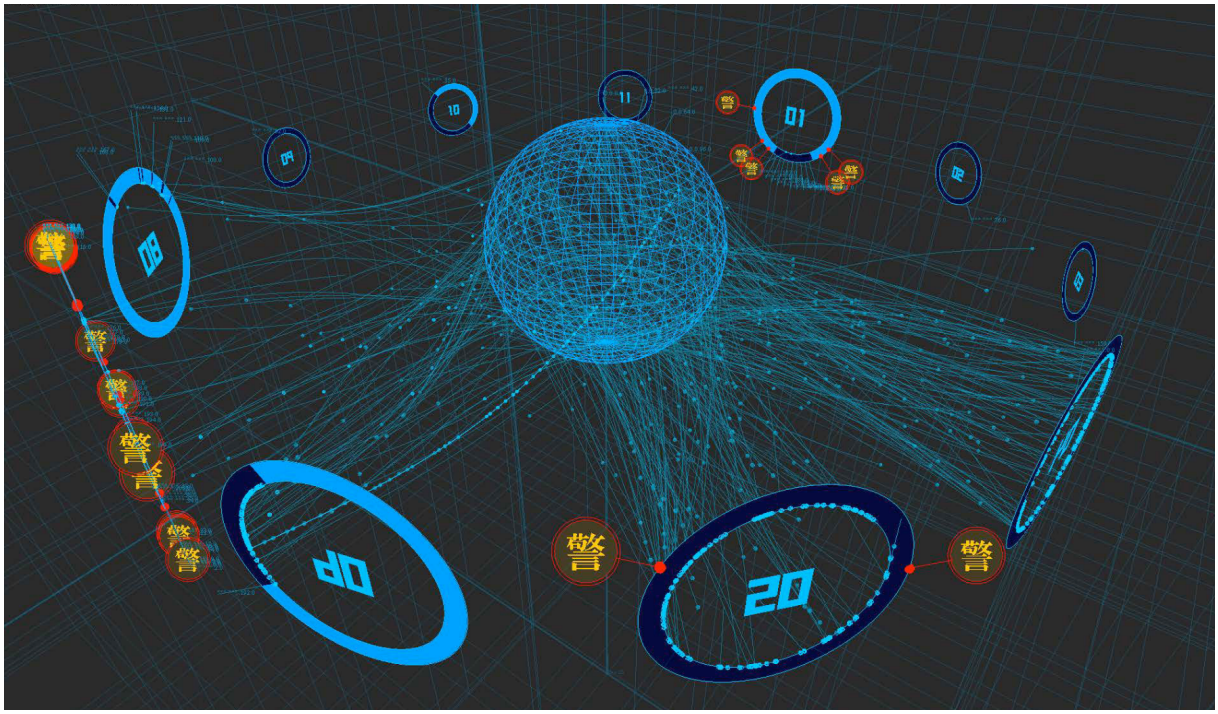


FIGURE 3.6 – DAEDALUS-VIZ [93] (2012, ©ACM) est un outil interactif permettant de surveiller en temps réel les adresses IP du darknet.

Des représentations 3D non immersives sont également proposées pour comprendre les événements sur un système à grande échelle. C'est le cas de DAEDALUS-VIZ [93] qui aide

à comprendre les alertes sur le darkweb grâce à une vue d'ensemble de la communication des différents serveurs basée sur une métaphore de système spatial en 3D (Figure 3.6). D'autres comme CyCop [9] sont utilisées dans le domaine du commandement cyber, principalement pour superposer des informations sur la carte du monde, mais ces méthodes sont davantage utilisées par les responsables que par les techniciens cyber.

Pour synthétiser, la recherche en Immersive Analytics pour les SOCs se concentre sur le challenge de la collaboration. Pour cela, elle utilise des visualisations 3D basées sur des augmentations des visualisations 2D pour les réseaux. Les résultats sont encourageants grâce à la bonne utilisabilité des prototypes testés. De plus, il semble que l'immersion permette d'améliorer la CSA partagée des participants. Cependant, ces résultats sont à contrebalancer par le faible nombre d'études et par le manque de comparaison avec des prototypes 2D.

3.3.3 Apports potentiels de l'Immersive Analytics pour la cybersécurité

L'Immersive Analytics semble adapté pour améliorer les visualisations des SOCs, mais les expériences comparant des visualisations immersives à des visualisations 2D de l'état de l'art restent peu nombreuses. Dans cette section, nous explorons les avantages que pourraient apporter l'Immersive Analytics aux visualisations des SOCs en nous basant sur des travaux connexes.

Comportement nominal

Nous avons montré dans le chapitre 2 que le temps et la comparaison sont des facteurs cruciaux dans le processus de visualisation pour la cybersécurité. Représenter des jeux de données complexes au cours du temps en 2D requiert de multiples vues, ce qui entraîne le besoin d'interactions qui peuvent surcharger l'opérateur. C'est particulièrement le cas pour les deux premières étapes du processus de réponse à alertes : l'observation des métriques puis la détection d'anomalies dans des réseaux.

Métriques Comme dit dans le chapitre précédent, l'opérateur commence son enquête sur les alarmes en regardant les métriques associées au système qu'il surveille dans le but d'identifier son comportement nominal et choisir une anomalie qui mérite d'être analysée. Pour représenter ces métriques et alertes au cours du temps, les visualisations 2D utilisent

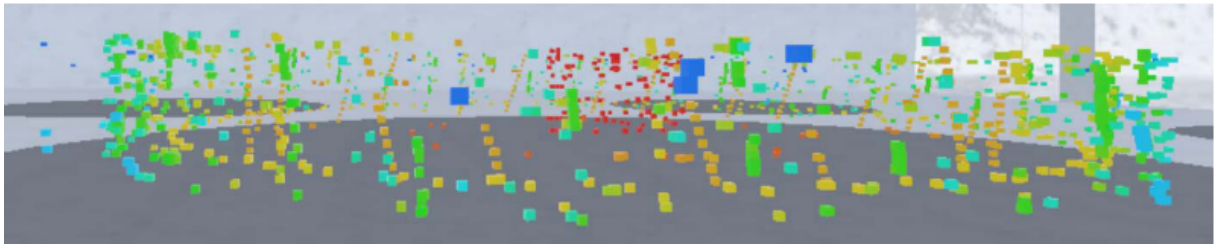


FIGURE 3.7 – Helovis [94] (2023, ©IEEE) utilise une représentation hélicoïdale de signaux électromagnétiques colorés en fonction des valeurs de leur fréquence pour créer des alignements indiquant la présence de signaux radar

majoritairement des histogrammes, qui ne mettent pas en valeur ces motifs périodiques. Des visualisations cycliques sont aussi utilisées, car elles ont l'avantage de mettre en valeur les motifs périodiques au prix d'un plus grand espace utilisé et d'une plus grande difficulté à les relier aux autres vues. Une représentation 3D qui permet de mettre en évidence un signal périodique est l'hélice ([95]). Elle a l'avantage de permettre de représenter les données dans l'espace qu'elle laisse disponible. Elle a été utilisée avec succès dans le domaine de la cybersécurité par Scott et al. en 2003 ([96]) qui ont utilisé des retours haptiques en plus d'un tableau de bord avec une représentation 3D pour représenter des réseaux soumis à différentes attaques. Cette expérience ne visait pas à comparer une représentation 2D avec une représentation 3D, mais plutôt à comparer une représentation 3D améliorée par des retours haptiques à la même représentation 3D sans retour haptique. Tant ([95]) que ([96]) ont proposé d'utiliser une interface avec un écran, un clavier et une souris, et non une solution immersive, car ces technologies n'étaient pas répandues à l'époque. Plus récemment, une représentation immersive de l'hélice (Figure 3.7) a été utilisée pour la guerre électronique et s'est avérée meilleure que les outils 2D ([94]). Cependant, les données utilisées dans cette expérience ne sont pas des données temporelles continues comme celles que l'on peut retrouver en cybersécurité ([81]). En outre, bien que la tâche consistait à détecter des motifs périodiques, il n'y avait pas d'objectif de corréler ces motifs à des alertes.

Réseaux La façon la plus courante d'afficher les réseaux en Immersive Analytics est d'utiliser des diagrammes nœuds/lien, les nœuds étant généralement représentés par des glyphes et les arêtes par des lignes entre ces glyphes. Des informations supplémentaires (telles que le type de données envoyées ou le statut d'un nœud) peuvent être transmises en utilisant la taille, la forme, la transparence et la couleur des nœuds, ainsi que la

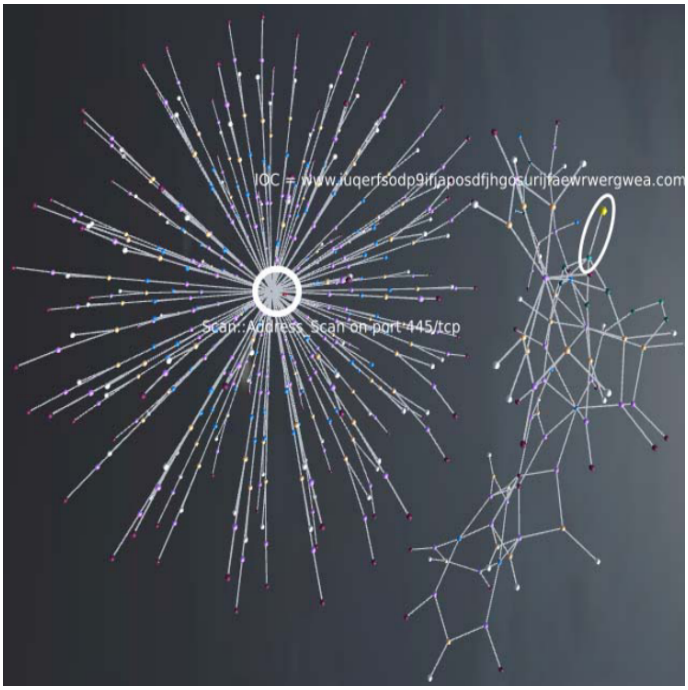


FIGURE 3.8 – Starlord [97] (2017, ©IEEE) utilise des graphs 3D pour représenter les clusters au sein d'un réseau

longueur, la forme, la transparence et la couleur des lignes. Différents algorithmes de génération de force sont utilisés pour limiter l'occlusion dans ces représentations. Il a été prouvé à plusieurs reprises que les représentations de diagrammes nœuds/liens sont plus efficaces en 3D qu'en 2D [78][98][29]. En effet, une solution immersive (qui permettait le suivi de la tête avec différents degrés de libertés de translation pour obtenir de la parallaxe de mouvement) permet d'afficher des réseaux 3 fois plus larges qu'en 2D sans perte d'information. Ces expériences restent cohérentes dans le temps (de 1998 à 2008) mais à cause des limitations des systèmes utilisés à cette époque, affichent un nombre de nœuds faibles (entre 32 et 1000) [81]. Les représentations graphiques en 3D sont utilisées dans deux articles de notre corpus : Niva [96]et Tudumi [99]. Dans le travail de Niva, la visualisation en 3D est combinée à un retour haptique qui illustre le danger représenté par l'alerte, tandis que Tudumi utilise la troisième dimension pour représenter le niveau d'accès des nœuds communicants. Un autre exemple est la visualisation graphique en 3D de Starlord (Figure 3.8) proposée par Leichtman et Al. [97], qui, bien qu'elle ne soit pas axée sur les alertes, reconnaît la nécessité de la 3D pour améliorer la compréhensibilité des diagrammes noeuds/lien de force dans le contexte des attaques de réseau.

Butscher et al. [100] ont introduit des représentations similaires à des coordonnées

parallèles en Réalité Augmentée (Figure 3.16). Dans cette approche, la troisième dimension est exploitée pour représenter le temps en plus des autres attributs sur les axes des coordonnées parallèles. Ces représentations sont associées à une table tactile à partir de laquelle les utilisateurs peuvent examiner des informations supplémentaires. Leur système s'est avéré efficace et a été bien accueilli par les utilisateurs qui devaient effectuer une tâche de détection de valeurs aberrantes dans les tendances d'un ensemble de données sociologiques.

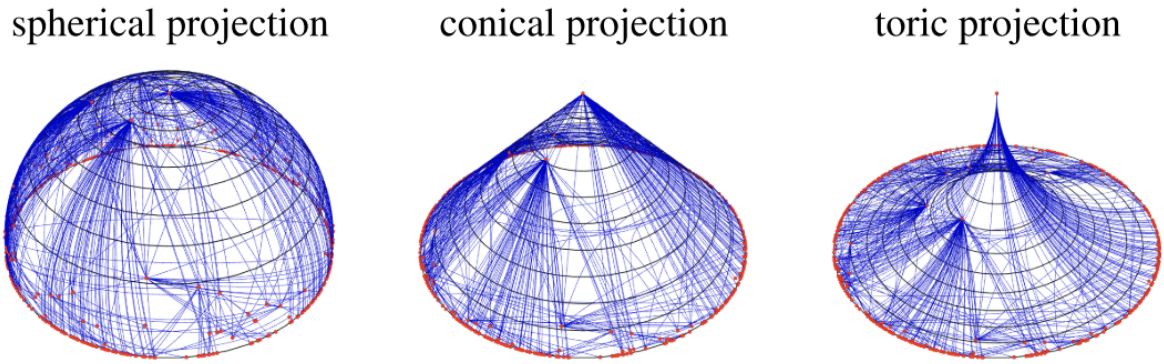


FIGURE 3.9 – Projections de réseaux sur différentes surfaces par Kobina et al. [101] (2022, ©HAL).

Pour mettre en avant certains attributs des réseaux, plusieurs auteurs proposent d'apporter des modifications aux graphes, par exemple en mettant en avant des nœuds spécifiques du réseau. Ainsi, une métaphore héliocentrique [102] d'un nœud permet de le représenter avec toutes les connexions qui apparaissent au fil du temps, ce qui permet à l'utilisateur de suivre les évènements du réseau sur une large fenêtre temporelle dans la même vue. Kobina et al. [101] proposent des systèmes de visualisations qui projettent des graphes 2D sur des surfaces 3D telles qu'une demi-sphère, un cône et une portion de tore (Figure 3.9). Le passage à la 3D améliore la visualisation de données complexes, en réduisant les chevauchements et en améliorant la perception de la connectivité. Les résultats comparatifs montrent que les présentations en 3D sont plus efficaces pour les tâches liées aux nœuds centraux et périphériques.

Des métaphores plus "réalistes" peuvent également être utilisées pour donner aux utilisateurs non experts une idée des événements qui se produisent sur le réseau. La métaphore de la ville est souvent utilisée pour représenter l'état d'un réseau. Dans le contexte de la cybersécurité, Carroll et Al. [103] utilisent une métaphore de feu pour

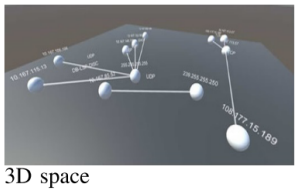
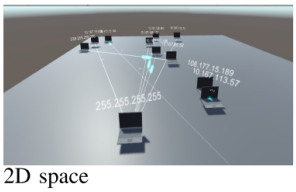
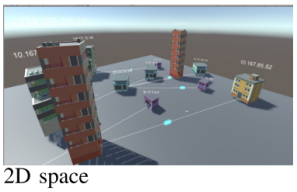
Abstract	Realistic	Metaphoric
		
3D space	2D space	2D space
Edges and nodes	Laptops and cables	Buildings and roads
Absolute positioning	Relative positioning	Relative positioning
Light	Packets	Cars

FIGURE 3.10 – Différents niveaux de métaphores pour visualiser un réseau par Carrol et al. [103] (2019, ©IEEE).

représenter un nœud attaqué (Figure 3.10). Toutefois, ces représentations n'ont pas été comparées à une représentation en 2D.

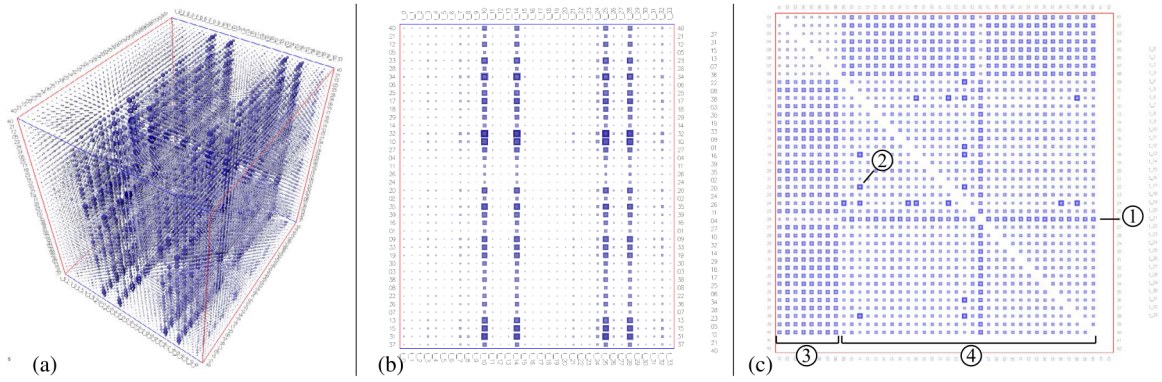


FIGURE 3.11 – Cubix [104] (2014, ©ACM) utilise une métaphore de matrice cubique pour représenter un réseau d'antennes au cours du temps et détecter des outliers.

Bien que les diagrammes nœuds/liens soient utiles pour mettre en valeur les clusters de communications au sein des réseaux [3], ils n'ont souvent pas la capacité de visualiser efficacement la dimension temporelle des données. À cette fin, le concept de matrices cubiques généralisé par [105] semble intéressant pour représenter les connexions d'un réseau tout en mettant en avant les changements au cours du temps. On peut voir la construction d'une matrice cubique dans le cadre de la cybersécurité comme une surimposition de matrices d'adjacence de réseaux sur différents pas de temps, formant ainsi une matrice temporelle. Inspirées par la compréhension humaine des cubes physiques, les matrices cubiques et leurs interactions constituent un moyen efficace pour les experts de naviguer à l'intérieur des données des réseaux [105] et d'interagir avec elles. Des implémentations no-

tables, comme Cubix[104] (Figure 3.11), ont démontré l'utilité des matrices cubiques par le biais de la collaboration avec des experts de différents domaines scientifiques. En outre, dans le cadre de la cybersécurité, cette approche a été appliquée à la tâche de surveillance des connexions à un serveur racine du DNS [106]. Bien que cette visualisation n'utilise pas de solution immersive, les matrices cubiques semblent être une solution prometteuse pour représenter les communications au sein d'un réseau au cours du temps en Immersive Analytics.

En résumé, la représentation classique sous forme de diagramme nœuds/liens pour les réseaux est plus efficace avec des systèmes immersifs qu'avec des systèmes 2D. Cependant, les études qui ont montré ce phénomène utilisent des systèmes dépassés et n'ont pas comme cas d'application la cybersécurité. Pour représenter le comportement du réseau au cours du temps, des visualisations basées sur des systèmes héliocentriques ou des matrices cubiques sont utilisées en cybersécurité, mais n'utilisent pas de systèmes immersifs et n'ont pas été comparés à des systèmes 2D. Ainsi, il semble qu'utiliser des visualisations 3D immersives des réseaux permettent résoudre les problèmes de représentations temporelles des visualisations 2D, bien que d'autres évaluations doivent encore être menées.

3.3.4 Space to Think

Nous avons vu dans le chapitre précédent des propositions d'organisation de la pensée des opérateurs grâce aux positions des visualisations dans l'espace 2D pour refléter leur cheminement intellectuel. Cette façon de faire rappelle la théorie de "l'espace pour penser" (Space To Think), qui souligne le rôle significatif que l'espace physique joue dans la génération de connaissances humaines [107]. Des études ont révélé que l'espace physique facilite une forme de cognition distribuée, où les analystes externalisent leurs processus cognitifs en organisant spatialement des documents et des visualisations dans l'espace [108]. Cette organisation spatiale améliore la synthèse des informations, ce qui améliore la compréhension des données par les opérateurs [107]. De plus, la capacité à naviguer physiquement dans cet espace favorise un accès et un rappel efficace des informations [80].

Cette théorie revêt une importance cruciale pour comprendre comment la spatialisation de l'information peut être utilisée pour améliorer la conception des systèmes d'analyse immersive. Les environnements virtuels riches en informations (IRVE) explorent comment les environnements virtuels peuvent représenter des propriétés et des attributs organisés dans l'espace afin de représenter des données plus efficacement que des visualisations classiques [80]. Grâce à l'intégration de la théorie de « l'espace pour penser » dans les IRVE,

FIGURE 3.12 – En haut, l'organisation de données géographiques par Satriadi et Al. [75] (2020, ©ACM). En bas, l'expérience de Lisle et Al. [107] (2021, ©IEEE) permet de trier des documents dans un environnement virtuel.

nous pouvons créer des systèmes d'analyse immersive qui permettent aux opérateurs cyber d'externaliser et d'organiser efficacement leurs processus cognitifs dans l'environnement, conduisant à une génération de connaissances et à une prise de décision plus efficaces.

En organisant leurs données dans l'espace, les opérateurs rendent leur cognition partiellement visible (Figure 3.12), fournissant des indices sur les activités de synthèse qui se déroulent au sein de leur processus cognitif. Cette visibilité peut être exploitée pour le développement d'algorithmes d'apprentissage automatique qui apprennent de l'activité cognitive de l'analyste et y répondent, améliorant ainsi l'efficacité du processus analytique. La mise en œuvre de cette théorie dans des systèmes tels que ForceSpire et StarSpire [109] prend en charge des scénarios d'analyse de texte sur de grands écrans, permettant aux utilisateurs d'organiser des documents, tandis que le système apprend et synthétise des documents pertinents dans l'espace structuré de l'utilisateur. Ces types d'interactions pourraient permettre, avec les algorithmes d'apprentissage automatiques de la cybersécurité, d'apprendre à classifier plus précisément des alertes en « observant » les raisonnements des opérateurs.

Corrélations des vues et interactions

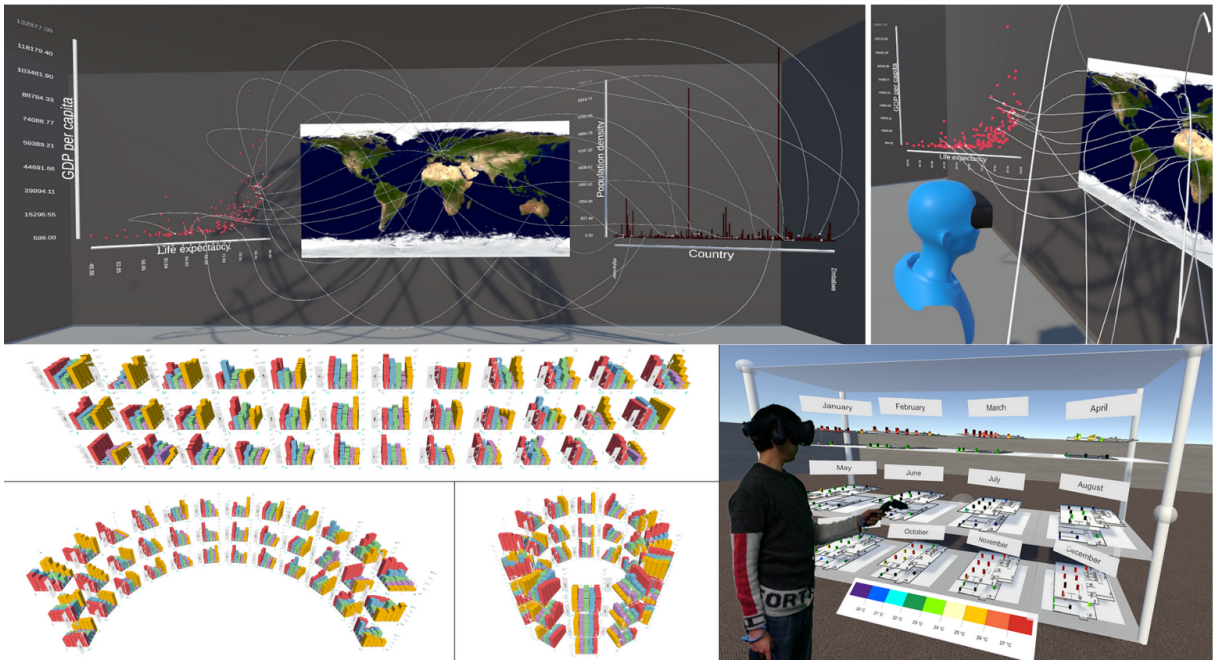


FIGURE 3.13 – En haut, Prouzeau et Al. [110] (2017, ©ACM) proposent un algorithme de création de lien entre visualisations pour minimiser l'occlusion. En bas, Liu et Al. [111] (2021, ©IEEE) étudient les différentes façons d'organiser des *small-multiples*.

Représenter les données en utilisant l'espace peut permettre de faciliter le travail des opérateurs, mais représente un challenge quand il s'agit de lier les différentes visualisations entre elles. Nous avons vu dans le chapitre précédent que les visualisations pour la cybersécurité utilisent des multivues pour représenter les différentes facettes des jeux de données. Cela entraîne l'usage de différentes techniques de visualisations et d'interactions pour maintenir la liaison entre les différentes vues. Dans le but d'utiliser un espace immersif pour représenter les données de cybersécurité, les mêmes questions de coordination entre les différentes vues et leur agencement se posent. En réalité virtuelle, le concept de small-multiple (Figure 3.13), permet de représenter plusieurs visualisations du même type et de les agencer en fonction des préférences de l'utilisateur [111]. De manière générale, les études sur les small-multiples montrent qu'il n'y a pas de préférence des utilisateurs et qu'ils vont suivre les algorithmes de recommandations [75], [112], ce qui est un avantage pour la conception de système de visualisation immersif, car cela permet d'organiser les vues entre elle de façon optimale pour corréler les données en suivant les bonnes pratiques de visualisations vues au chapitre précédent. En ce qui concerne les liens entre les

différentes visualisations, Prouzeau et al. [110] présentent un algorithme pour optimiser la représentation des liens entre différentes visualisations pour minimiser l'occultation et l'encombrement dans des environnements immersifs collaboratifs. En somme, les environnements 3D immersifs permettent la représentation et la corrélation des données et des vues de façons novatrices et probablement plus efficaces que la 2D, mais il manque des évaluations comparatives pour le prouver.

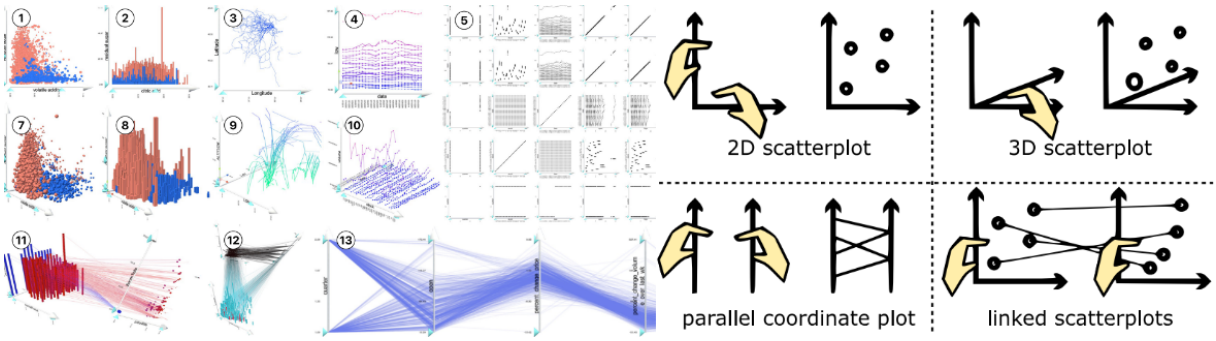


FIGURE 3.14 – À gauche, les différentes possibilités de visualisations de la librairie IATK [113] (2019, ©IEEE). À droite, les interactions proposées dans ImAxes [112] (2020, ©IEEE).

Ainsi, la recherche en Immersive analytics a proposé des solutions pour la coordination des multiples vues dans un environnement immersif, mais comme en 2D, l'usage de multiples vues demande souvent des interactions avec celles-ci (Figure 3.14). Pour lier plusieurs vues entre elles, les solutions d'Immersive Analytics se concentrent majoritairement sur une utilisation des contrôleurs fournis par les fabricants de casques, par exemple, la librairie IATK [113] intègre une liaison des vues grâce à une technique de *brush and highlight*. En revanche, pour d'autres types d'interactions, des interfaces tactiles ou tangibles ont été développées pour permettre la manipulation des différentes vues dans l'environnement immersif, ainsi que la sélection et l'annotation des données, ce qui correspond à certaines des interactions les plus utilisées en cybersécurité. En revanche, la manipulation de widgets qui permettent de filtrer et d'explorer les données plus en détail reste encore globalement inexplorée, ce qui peut être un inconvénient des visualisations pour la cybersécurité, car la complexité de données peut nécessiter des fonctions de filtrage avancé. De plus, nous avions vu que les interactions abstraites, comme modifier des valeurs et appliquer les filtres en utilisant boutons ou des requêtes, sont les interactions majoritaires dans les visualisations pour la cybersécurité. Ce type d'opération nécessite un accès aux outils classiques de la cybersécurité dans l'environnement virtuel, qui bien que possible

reste actuellement limité [114].

3.3.5 Espace de Conception

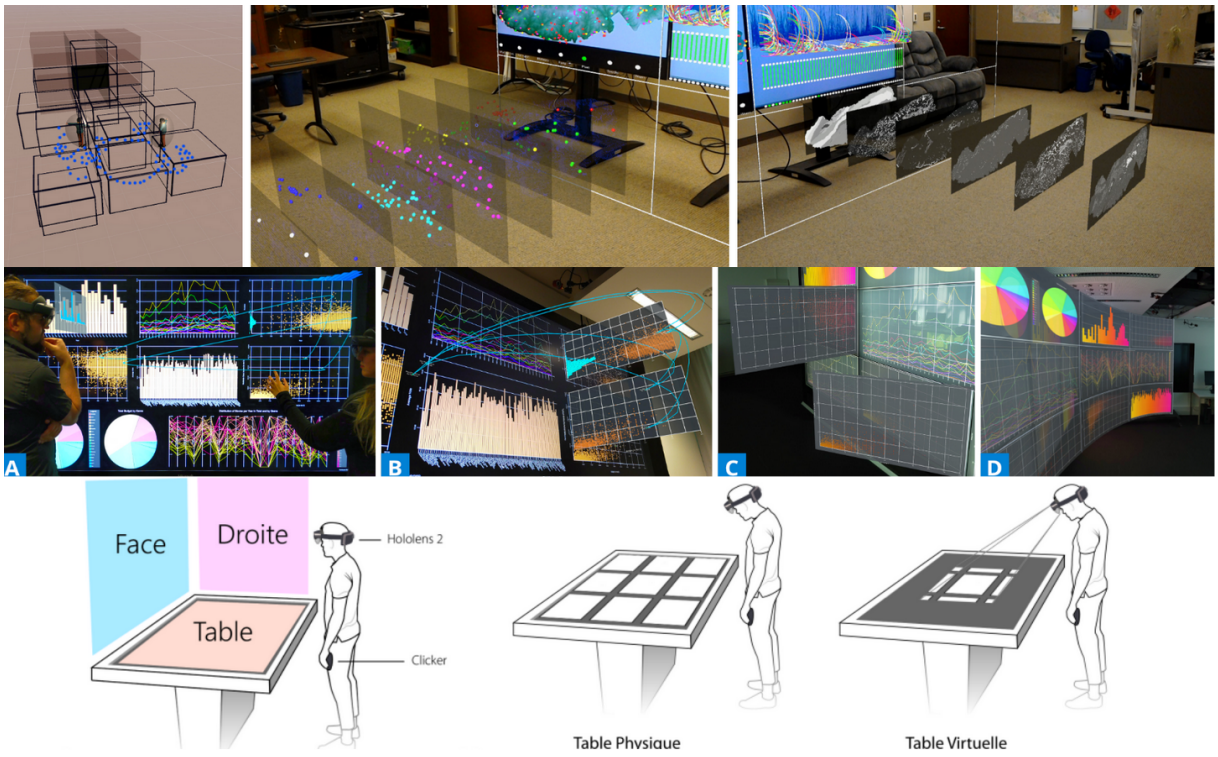


FIGURE 3.15 – En haut, Mahmood et Al. [115] (2018, ©IEEE) proposent un système de coordinations de multiples vues 2D. Au milieu, Reipschlager et Al. [116] proposent un Espace de Conception de l'augmentation d'affichage sur écran. En bas, Perelman et Al. [117] (2022, ©IEEE) ont étudié les transitions visuelles entre un affichage horizontal sur une table et deux affichages virtuels verticaux

Ces techniques de visualisations et d'interactions forment l'espace de conception de l'organisation d'interfaces multivues en Immersive Analytics. Plusieurs études ont exploré cet espace de conception (Figure 3.15), surtout en Réalité Augmentée. Mahmood et al. [115] proposent une librairie permettant de créer de Multiples Vues Coordonnées (MVC), une version en 3D des multivues, intégrant diverses visualisations 2D adaptées aux données représentées dans un espace de travail d'analyse en Réalité Augmentée. Cette approche combine des dispositifs de Réalité Augmentée avec des visualisations 2D classiques sur grand écran. Dans la continuité de ce travail, Liu et al. [111] présentent un espace de conception pour améliorer la visualisation des données multivariées sur de grands écrans

interactifs et augmentés par de la RA, en présentant un espace de conception systématique englobant l'alignement spatial, la visualisation augmentée et la fourniture de contenu personnel. Tout en proposant un espace de conception de l'Immersive Analytics pour l'augmentation de visualisations sur grand écran, il reconnaît l'absence de bonnes pratiques pour choisir les visualisations en fonction des données et des tâches, en particulier en ce qui concerne l'alignement spatial. Afin de contribuer à la création de bonnes pratiques pour l'organisation des vues dans un espace 3D, Perelman et Al. [117] étudient les transitions visuelles entre des données présentes sur une table horizontale et des affichages verticaux en Réalité Augmentée afin d'effectuer des recommandations à propos de l'affichage optimal des données. Enfin, Ma et Millet [118] fournissent des recommandations sur la conception de dashboard immersif ainsi que sur les façons de les évaluer. Ils insistent entre autres sur la nécessité de permettre aux utilisateurs d'avoir le contrôle sur tout le pipeline de visualisation au sein de l'environnement immersif. En revanche, ces recommandations ne s'étendent pas sur le choix spécifique de visualisation ou d'interactions en fonction du but de l'utilisateur et du type de données. Ainsi, la recherche en Immersive Analytics s'est concentrée sur le fait de fournir un Espace de Conception des visualisations, mais ne réussit pas encore à fournir des recommandations pour obtenir des représentations les plus efficaces possibles.

3.3.6 Collaboration

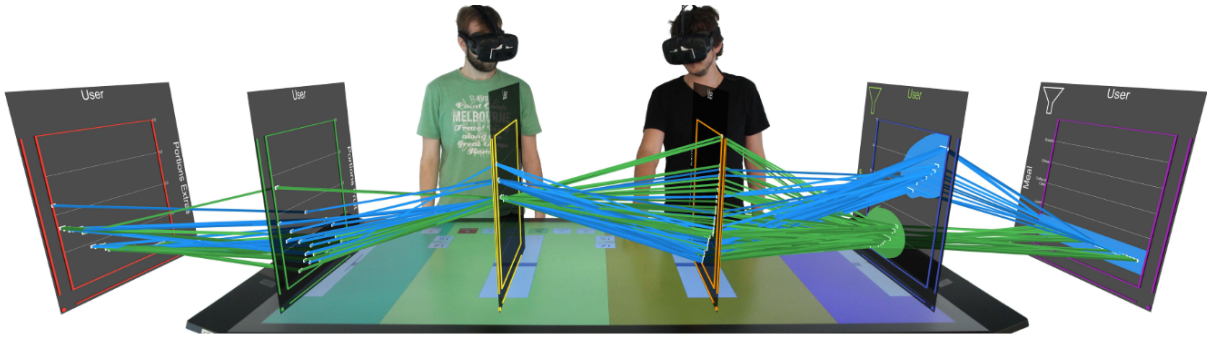


FIGURE 3.16 – L'outil proposé par Butscher et al. [100] (2018, ©ACM) permet de faciliter l'analyse collaborative de données multidimensionnelles.

Dans la section concernant l'usage de l'Immersive Analytics pour la cybersécurité, nous avons vu que de manière générale, les environnements virtuels collaboratifs permettent de partager les informations émotionnelles et physiologiques des opérateurs, ce

qui permet d'améliorer la collaboration au sein d'une équipe. En particulier, ils permettent des interactions interhumaines au sein de l'environnement immersif qui rendent plus efficace le partage d'informations [119]. Par exemple, Lee et Al. [108] proposent Fiesta un prototype d'Immersive Analytics qui permet de positionner librement des interfaces de visualisation 2D ou 3D n'importe où dans un environnement virtuel collaboratif. Ainsi, il semble logique que la recherche en visualisation pour la cybersécurité se soit emparé de l'Immersive Analytics pour améliorer la collaboration au sein des équipes des SOCs. Comme nous l'avons dit dans la section 3.3.2, les premiers résultats semblent indiquer une meilleure collaboration lors de l'usage d'outils immersifs, mais cela reste à confirmer par des expériences supplémentaires.

3.3.7 Limitations

FIGURE 3.17 – Intégration des interfaces WIMP dans l'environnement. À gauche, l'outil uDesktopDuplication [3]d'hecomi, à droite le prototype proposé par Hoppe et Al. [114] (2020, ©Springer).

Les limitations des espaces de travail de l'analyse immersive résident dans les défis liés à l'intégration des fonctionnalités existantes des applications 2D dans l'espace 3D. Par exemple, le chargement de données et la modification de données sont des tâches courantes en cybersécurité, mais sont rarement réalisées par les participants des expériences d'Immersive Analytics. Souvent, les données sont préchargées avant les expériences. Cette méthode fonctionne dans le cadre d'une expérience scientifique, mais ne convient pas au passage en production. Il existe des solutions d'intégrations des interfaces WIMP (Windows, Icons, Menus and Pointing device pour « fenêtres, icônes, menus et dispositif de pointage ») dans l'environnement virtuel [114], ce qui pourrait permettre de résoudre aussi bien les problèmes d'accès aux outils classiques des opérateurs qui permettent de charger les données que des interactions abstraites. En revanche, ces intégrations de fenêtres

Challenges de l'Immersive Analytics		Immersive Analytics pour la cybersécurité					
		Travaux actuels	Métriques	Réseaux	Espace à penser	Corrélations	Limitations
Nouveaux systèmes immersifs	Multimodalité		+	+		+++	
	Visualisations	✔	+++	+		+	
	Interactions	✔	+	+++	+++	+++	✘
	Collaboration	✔✔✔			+		
Evaluation	Immersives vs 2D	✔	+++	+++	+++	+++	
	Immersives vs Immersives	✔✔✔	+	+++		+++	
	Nouvelles Modalités d'évaluation		+	+	+++	+	
Design Space et Bonne Pratiques			+++	+++	+++	+++	
Mise en production	Integration au workflow existant				+++	+	✘✘✘
	Librairies				+++		✘

TABLE 3.1 – Ce tableau illustre les relations entre les travaux actuels, les opportunités et les limites de l'Immersive Analytics pour la cybersécurité, avec les challenges de l'Immersive Analytics. Le nombre de symboles reflète la contribution du défi des visualisations pour la cybersécurité au défi correspondant de l'Immersive Analytics. Un symbole indique des contributions mineures, trois symboles indiquent des contributions majeures.

dans l'environnement immersif ne visaient pas l'Immersive Analytics, et il n'existe pas à notre connaissance de travaux sur l'intégration de ces fenêtres dans le workflow d'un opérateur. De plus, les interactions avec les fenêtres se font souvent avec des méthodes de raycasting [114] (Figure 3.17) ce qui n'est pas optimal pour interagir avec des fenêtres 2D . À notre connaissance, il existe peu de recherches sur l'intégration de ces interfaces dans l'environnement immersif, surtout vis-à-vis du changement de paradigme entre le classique clavier/souris pour interagir avec les fenêtres 2D et les interfaces immersives pour interagir avec les objets 3D.

3.3.8 Conclusion

La recherche en Immersive Analytics explore l'utilisation des technologies immersives pour une réponse collaborative aux alertes de cybersécurité, mettant l'accent sur la compréhension des topologies réseau, la contextualisation des alertes d'intrusion, et l'amélioration de la Common Cyber Situational Awareness. Ainsi, les recherches en Immersive Analytics pour la cybersécurité répondent au challenge de la création d'interface de visualisation immersif et collaboratif avec des résultats encourageants, mais il reste encore à poursuivre les évaluations par rapport à des visualisations 2D. D'autres solutions déjà proposées par la communauté de l'Immersive Analytics pourraient également être utiles aux visualisations pour la cybersécurité. Dans cette section, nous lions les limites des

3. https ://github.com/hecomi/uDesktopDuplication, 02/10/2023

représentations pour la cybersécurité identifiées au chapitre 2 aux solutions que peut apporter l'Immersion Analytics et aux challenges que cela représente du point de vue de l'Immersive Analytics.

Le premier point d'amélioration est l'affichage des métriques des systèmes observés. Celles-ci sont généralement affichées avec des histogrammes 2D qui sont facilement reconnaissables par les opérateurs cyber mais ne disposent pas d'avantages particuliers, surtout pour mettre en valeur des comportements nominaux et des comportements suspects, c'est-à-dire des motifs périodiques et des potentielles ruptures. Afin de mettre en valeur ces motifs périodiques, l'utilisation de représentations cycliques a été un peu explorée, mais l'avantage qu'elles ont de pouvoir mettre en valeur les signaux périodiques vient avec les inconvénients du manque de place et de la difficulté de la corrélation avec les autres données. Pour pallier cela, l'utilisation d'hélices, des représentations 3D qui reprennent le concept d'alignement des motifs des représentations cycliques, mais l'étendent à la troisième dimension, est une piste de recherche prometteuse. Ces hélices ont été utilisées dans le cadre de visualisations cyber mais les recherches récentes dans des environnements immersifs ont montré leur efficacité par rapport à des visualisations 2D, car on peut alors utiliser notamment l'espace disponible qu'elles laissent pour les corréler à d'autres données. Ainsi, il est pertinent d'utiliser les hélices immersives en cybersécurité pour représenter les données pouvant contenir des motifs périodiques et les lier à d'autres données. La représentation des métriques en cybersécurité correspond principalement aux challenges de trouver de nouvelles visualisations immersives et de les évaluer par rapport aux visualisations 2D existantes. Dans une moindre mesure, ces évaluations auront de nouvelles modalités et apporteront de nouvelles interactions qu'il conviendra d'évaluer.

Le deuxième point concerne la représentation des réseaux, car les représentations 2D (graphes/matrices/coordonnées parallèles) utilisées en cybersécurité sont limitées pour représenter le passage du temps. Pour le faire, elles utilisent des glyphes et des méthodes d'animations pour représenter les variations des données au cours du temps. Ces méthodes demandent beaucoup d'efforts de la part de l'opérateur. Les représentations immersives 3D permettent d'améliorer les performances des visualisations des réseaux, en particulier au cours du temps. Premièrement, les graphes 3D permettent de mieux visualiser les relations au sein des réseaux. Des représentations 3D immersives ont déjà été évaluées en dehors de la cybersécurité, et l'Immersive Analytics pour la cybersécurité a déjà utilisé ces types de représentations pour visualiser des topologies de grands réseaux. Deuxièmement, la troisième dimension permet de représenter au cours du temps l'état des réseaux, avec

des matrices cubiques ou des coordonnées parallèles 3D. En revanche, seules les matrices cubiques ont été implémentées dans un cadre de cybersécurité, mais n'ont pas été évaluées face à des représentations 2D. De plus, les représentations implémentées n'ont pas proposé d'interactions immersives novatrices par rapport à l'existant, ce qui peut être une piste d'amélioration. Ainsi, la représentation immersive des réseaux en cybersécurité correspond aux challenges d'évaluation des représentations immersives non seulement entre elles, mais aussi par rapport aux visualisations 2D. De nouvelles modalités d'évaluations pourraient être intéressantes à explorer.

Le troisième point concerne le concept d'espace à penser, celui-ci a été exploré par la communauté cyber et rencontre son équivalent dans la communauté d'Immersive Analytics. Le passage à la 3D immersive, en plus d'ajouter une dimension pour organiser l'espace, permet d'offrir de nouvelles interactions. L'utilisation de l'espace à penser en cybersécurité facilite la collaboration, améliore la gestion d'enquêtes complexes et entraîne les algorithmes de détection d'intrusion grâce à l'organisation spatiale des données. Ainsi, l'utilisation d'espace à penser immersif en cybersécurité correspond aux challenges d'interactions, de collaborations, et appelle à de nouvelles modalités d'évaluations pour pouvoir comparer les espaces à penser immersifs par rapport aux espaces à penser 2D. De plus, ces nouveaux espaces immersifs nécessitent des libraires de visualisations 3D pour être réalisés, et doivent être évalués par rapport à leur capacité de s'intégrer au workflow des opérateurs cyber pour qu'ils soient acceptés en production.

Le quatrième point concerne l'organisation des visualisations au sein des environnements virtuels. Il existe des propositions d'espace de conception de l'Immersive Analytics qui explorent différentes manières de lier les données entre elles pour fluidifier le processus de visualisation, que ce soit en les organisant par rapport à un axe commun, avec des liens, ou encore avec des méthodes classiques de mise en valeur. De même, la communauté de l'Immersive Analytics a proposé de nombreuses interactions avec les systèmes immersifs. Cependant, il manque des évaluations de cet espace de conception pour savoir quelle est la meilleure façon d'afficher et de corréler ces données afin de fournir des lignes directrices de l'Immersive Analytics. De plus, les interactions proposées permettent difficilement d'utiliser des widgets ou de réaliser des interactions abstraites, nécessaires en cybersécurité. Ainsi, l'organisation de l'espace et les interactions correspondent aux challenges de l'Immersive Analytics, de la recherche de nouvelles représentations et d'interactions multimodales, ainsi que de l'évaluation des représentations immersives.

La principale limitation de l'adoption de l'Immersive Analytics est son incapacité ac-

tuelle à s'adapter au workflow des opérateurs cyber. En particulier, étant donné que les opérateurs n'apprécient pas particulièrement l'utilisation de visualisations, leur réticence à l'égard de la réalité virtuelle est encore plus prononcée. Les solutions actuellement proposées sont spécifiques à certaines situations et peu adaptables pour traiter la diversité des données en cybersécurité. Bien que des librairies de visualisation en Immersive Analytics existent, elles n'ont pas encore été exploitées dans le contexte de la conception automatique des visualisations. Actuellement, il n'est pas aisé d'interagir avec les outils de cybersécurité depuis un environnement virtuel, ce qui met en évidence le manque de recherches en Immersive Analytics visant à proposer des systèmes techniques permettant d'intégrer un workflow dans un environnement virtuel. Cette limitation correspond aux challenges d'adapter l'Immersive Analytics à un workflow spécifique.

Ainsi, dépasser les limites des visualisations de cybersécurité 2D à l'aide de l'Immersive Analytics contribue à l'avancée des recherches sur presque tous challenges de l'Immersive Analytics. En effet, au vu des problématiques que nous avons détectées en cybersécurité, seule la Situational Analytics ne semble pas pertinente pour les visualisations des SOCs.

Résumé

Dans ce chapitre, nous examinons comment l'Immersive Analytics peut potentiellement aider à surmonter les limites de visualisations 2D pour la cybersécurité identifiées au chapitre précédent. Les technologies de Réalité Virtuelle se sont répandues et ont amélioré leurs performances au cours des 10 dernières années. Depuis lors, le domaine de l'Immersive Analytics explore différentes façons d'utiliser ces technologies immersives pour représenter des données. Les principaux défis de l'Immersive Analytics sont la création de nouvelles interfaces multimodales pour la représentation des données, leur évaluation, et leur transfert industriel. Enfin, la collaboration est l'aspect qui a été le plus étudié dans la recherche en Immersive Analytics pour la cybersécurité, avec des évaluations prometteuses qui encouragent la poursuite des travaux. Nous avons observé le potentiel d'utilisation de la 3D pour représenter les données métriques et les réseaux au fil du temps. Premièrement, la recherche sur les comportements périodiques en 3D et en Immersive Analytics pour les données périodiques se concentre sur les hélices, mais n'a pas été évaluée dans un contexte similaire à celui de la cybersécurité, ni avec des prototypes correspondant aux enjeux des métriques en cybersécurité. Deuxièmement, la recherche a proposé et évalué de nombreuses façons de représenter les réseaux en 3D, qui se sont révélées plus efficaces que les représentations 2D. En ce qui concerne la représentation du temps, une piste prometteuse est l'utilisation de matrices cubiques, dont un prototype existe en cybersécurité. Cependant, il n'y a pas encore eu d'implémentation immersive ni d'évaluation. Le troisième point concerne l'utilisation des positions relatives des visualisations de données dans l'espace pour corréler différentes représentations. Quelques concepts de conception ont été proposés mais n'ont pas encore été évalués pour démontrer l'intérêt de l'utilisation de la 3D pour la corrélation des données. En plus de l'utilisation de l'espace 3D pour la corrélation des données, celui-ci peut permettre à un opérateur d'organiser un raisonnement complexe afin d'améliorer ses capacités de réflexion, ainsi que sa collaboration avec son équipe ou ses interactions avec des algorithmes intelligents. Nous constatons également certaines limites de l'Immersive Analytics, notamment le manque d'intégration des outils 2D dans l'environnement immersif pour un workflow fluide, ainsi que l'absence d'interactions dédiées à leur utilisations.

Chapitre 4

Cybercopter : Immersive Analytics pour la classification des alertes basée sur des données périodiques.

Nous avons vu dans les chapitres précédents l'importance de représenter les métriques qui caractérisent un système de façon à mettre en avant les tendances périodiques et les déviations de celles-ci, qui pourraient indiquer une attaque ou un dysfonctionnement. Les visualisations principalement utilisées ont des défauts pour arriver à cette fin. Les histogrammes ne permettent pas de mettre en valeurs les signaux périodiques, les représentations cycliques comme les spirales ou les cercles concentriques prennent beaucoup de place et rendent difficile la corrélation avec le reste des données. Dans ce chapitre, nous présentons l'état de l'art spécifique aux visualisations de ce type de données pour ce type de tâches : corréler des alertes à des ruptures de périodicité. Puis, nous nous intéressons aux visualisations qui répondent à des problèmes similaires en Immersive Analytics. Puis, après avoir constaté qu'aucune ne répondait exactement au problème que nous avons identifié : la représentation de données périodiques et leur corrélation à des alertes, nous présentons notre concept de visualisation et ses implémentations. Du point de vue de la cybersécurité, ce concept permet de représenter des métriques en mettant en valeurs des comportements périodiques tout en les corrélant avec d'autres données. Du point de vue de la visualisation des données, ce concept montre comment corréler des données temporelles périodiques à d'autres données temporelles en utilisant l'Immersive Analytics et les interactions associées. Ensuite, nous présentons deux mises en œuvre distinctes de ce concept, illustrant des jeux de données variés et recourant à différents dispositifs d'interaction pour manipuler les visualisations. Nous avons évalué ces prototypes auprès d'utilisateurs experts et novices en cybersécurité, démontrant l'utilité des hélices pour

corréler différents types de données à des motifs périodiques. De plus, ils ont montré une efficacité accrue dans la réponse aux alertes par rapport aux visualisations 2D de l'état de l'art. Les évaluations de ces prototypes nous permettent de formuler des recommandations sur l'utilisation des visualisations immersives en cybersécurité, en particulier en ce qui concerne les dispositifs d'interaction. Cependant, les experts ont noté que l'utilisation de notre prototype les isolait de leurs outils classiques, ce qui rejoint les limitations décrites dans le chapitre 3.

4.1 État de l'art sur les visualisations cycliques

Pour mettre en évidence les signaux périodiques dans les données affichées, certaines personnes utilisent des représentations classiques telles que les histogrammes pour afficher l'activité du réseau au fil du temps, comme Gove et al. ([25]). D'autres utilisent des représentations spécifiques pour cette tâche, comme les spirales ou les cercles concentriques. Les spirales interactives mettent en évidence des motifs dans les données temporelles en reliant des valeurs similaires selon les lois de Gestalt telles que la similitude et la continuité ([120]). Dans le domaine de la cybersécurité, ([121]) a utilisé ces spirales pour représenter les données de capteurs présentant des comportements périodiques.

Foresti ([122]), Legg ([123]) et Ngoc Anh Huynh ([124]) utilisent des cercles concentriques pour mettre en évidence des schémas périodiques dans un ensemble de données. Ils permettent de corréler les données représentées sur les cercles avec d'autres informations, telles que les lieux où les alertes sont apparues, en prévoyant suffisamment d'espace ([122]) au centre de la représentation. Cet avantage est atténué par l'impossibilité de modifier le pas de temps représenté. Ces visualisations ont également l'inconvénient d'exiger de grands espaces de visualisation, car elles doivent utiliser trois dimensions de l'espace de visualisation (deux pour afficher la série temporelle sous forme cyclique et la troisième pour afficher sa valeur) pour afficher une série temporelle bidimensionnelle.

En outre, pour relier des représentations en spirale ou circulaires à d'autres, on peut utiliser soit des liens qui passent au-dessus des représentations, soit le changement de couleur des représentations. En effet, leur géométrie empêche d'utiliser un axe commun avec une autre visualisation pour relier les données entre elles. Dans les deux cas, des informations sont perdues, soit à cause de l'occlusion, soit parce que l'utilisation de la couleur pour la corrélation est un inconvénient si elle est déjà utilisée pour représenter une dimension des données. Ce dernier inconvénient est particulièrement important pour les visualisa-

tions périodiques où la couleur est nécessaire pour représenter les séries temporelles 4.1. La difficulté de corréler ces représentations avec d'autres types de représentations peut constituer un problème pour la connaissance de la situation de l'opérateur, étant donné que différents types de visualisations doivent être utilisés pour afficher différents types de données.

Les environnements 3D offrent une autre dimension pour représenter les données, qui peut être utilisée pour corréler les représentations sans perdre d'informations. Par exemple, elle peut aider à relier des données sans occlusion ([110]) et à les relier avec des données abstraites en utilisant leur placement ([125]), ou des données réelles ([126]) telles que les positions des capteurs.

L'hélice est une représentation 3D qui permet de mettre en évidence un signal périodique ([95]). Elle a été utilisée avec succès dans le domaine de la cybersécurité par Scott et Al. en 2003 ([96]) qui ont utilisé des retours haptiques en plus d'un tableau de bord avec une représentation 3D pour représenter des réseaux soumis à différentes attaques. Cette expérience ne visait pas à comparer une représentation 2D avec une représentation 3D, mais plutôt à comparer une représentation 3D améliorée par des haptiques à la même représentation 3D sans haptique. Tant Gautier et Al. ([95]) que Scott et Al. ([96]) ont proposé d'utiliser une interface avec un écran, un clavier et une souris, et non une solution immersive, car les technologies immersives n'étaient pas répandues à l'époque. Plus récemment, une représentation immersive de l'hélice a été utilisée pour la guerre électronique et s'est avérée meilleure que les outils 2D ([94]). Cependant, les données utilisées dans cette expérience ne sont pas des données temporelles continues comme celles des capteurs SWaT ([81]). En outre, bien que la tâche consistait à détecter des motifs périodiques, il n'y avait pas d'objectif de corréler ces motifs à des alertes.

Pour évaluer les avantages de l'analyse immersive, des paramètres subjectifs tels que la facilité d'utilisation et la charge de travail sont souvent mesurés en plus des paramètres quantitatifs tels que le temps d'exécution. Nous soutenons que l'état de flow des participants devrait également être pris en compte. Le flow est un état psychologique dans lequel les individus sont totalement immergés dans une activité, perdant la notion du temps et de l'environnement, et qui est souvent associé à des performances élevées et à des émotions positives ([127]). Comme les opérateurs de cybersécurité doivent prendre des décisions rapides avec un minimum d'erreurs, l'amélioration de leur état de fluidité peut améliorer leur concentration et la réalisation de leurs tâches. Des recherches antérieures ont exploré la manière dont la réalité mixte peut induire un état de fluidité chez les

utilisateurs (**Kim2022**), et l'analyse immersive pourrait offrir des avantages similaires.

4.2 Concept

Dans cette section, nous présentons le concept de visualisation théorique. Nous utilisons des hélices pour représenter des motifs périodiques dans le même principe qu'avec les spirales, mais sans les inconvenant du manque de place et de difficulté à corréler les spirales avec d'autres visualisations.

4.2.1 Les limitations des spirales

Dans une spirale, le temps est courbé du centre vers l'extérieur, la couleur de chaque point de données correspond à la valeur du capteur (figure 4.1). Les points de la spirale qui traversent la même ligne droite passant par le centre de la spirale sont séparés par le même nombre de points de données, donc la même durée. La période de la spirale peut être définie par ce nombre. En modifiant la période, on peut faire apparaître des motifs périodiques dans les données affichées. En effet, si la période du motif est la même que la période de la spirale, les motifs s'aligneront radialement et seront facilement reconnaissables grâce à la loi de proximité de la gestalt. Bien que les spirales aient l'avantage de modifier la périodicité de leur représentation, ce qui est très utile lorsque l'on recherche des ruptures de périodicité, elles ont des limites pour les corrélations d'attributs. En effet, la meilleure façon de corréler deux attributs est de les représenter sur une échelle commune ([128]), ce qui n'est pas possible en 2D car les deux axes sont déjà utilisés pour représenter des spirales ou des cercles. C'est pourquoi les solutions 2D habituelles utilisent d'autres canaux, tels que la largeur ou la teinte, pour représenter les attributs sur la visualisation principale (sous-figures c et d sur figure 4.9). Toutefois, la combinaison du canal visuel de la largeur avec les canaux de la teinte et de la forme peut être source de confusion et d'inconfort, car ils ne sont pas entièrement séparables dans l'œil humain ([29]). Une autre solution consiste à utiliser une visualisation supplémentaire pour représenter des attributs additionnels et à relier les deux visualisations par des liens, ce qui peut être préjudiciable, car cela peut cacher ou occulter d'autres informations.

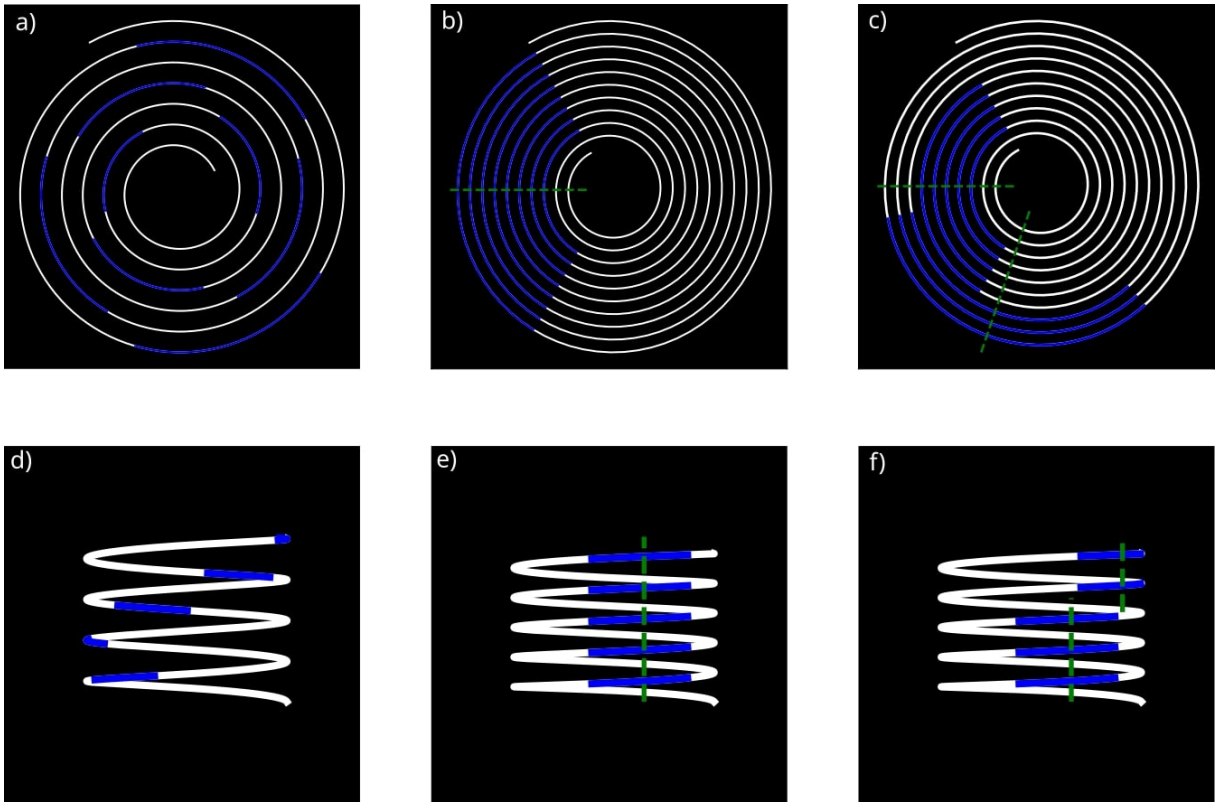


FIGURE 4.1 – Comparaison entre les spirales et les hélices, Le même signal est représenté sur les visualisations a) et b). Il est possible de faire apparaître des motifs périodiques en changeant la période de la spirale. Le même principe se produit sur les visualisations d) et e) en changeant la période de l'hélice. Sur les visualisations b) et e), tous les points de la spirale (ou de l'hélice) qui traversent la ligne verte sont espacés de la même période de temps. Sur les visualisations c) et f), une rupture dans le schéma périodique apparaît, les distances entre les deux lignes vertes représentent un déphasage.

4.2.2 Choix de l'hélice

Nous essayons de représenter les données de manière à souligner la périodicité de la même manière que les spirales : en modifiant la période de la représentation pour mettre en évidence les ruptures de périodicité. Cependant, la représentation doit également permettre de corréler les ruptures de périodicité avec les alertes. Pour ce faire, nous remplaçons la spirale par une hélice, ce qui permet d'ajouter un nouvel axe à la visualisation en utilisant la troisième dimension (figure 4.1).

La formule utilisée pour produire l'hélice est la suivante :

$$(y, r, \theta) = \begin{cases} t, \\ R, \\ -2\pi(t \bmod p) \end{cases}$$

où (y,r,θ) représente les coordonnées d'un point de l'hélice dans un système de coordonnées cylindrique , t la valeur du temps, R une constante qui définit la taille de l'hélice, et p la période de l'hélice

D'une façon similaire à celles des représentations cycliques 2D, les représentations hélicoïdales permettent de mettre en valeurs les motifs périodiques en les alignant. De plus, comme les points de l'hélice ne changent pas de position par rapport à l'axe vertical, cela permet de corréler d'autres représentations à l'hélice sans interférer avec la représentation des signaux périodiques. Cela peut permettre de comparer deux hélices avec des périodes différentes ou encore de corréler d'autres attributs à la représentation hélicoïdale. Enfin, une différence entre les hélices et les spirales est que chaque cercle à la même longueur dans une hélice alors qu'il s'agrandit dans une spirale, ce qui déforme la perception du temps.

4.3 Prototype I

Pour évaluer le concept de l'hélice, nous avons réalisé un premier prototype en collaboration avec les partenaires industriels de la chaire : le Cybercopter. Ce prototype a été réalisé et évalué lors des restrictions dû à la COVID19, ce qui nous a conduit à modifier l'idée initiale d'une visualisation immersive en une visualisation 3D pour pouvoir l'évaluer à distance. Nous avons testé son utilisabilité avec de bons résultats et des retours utilisateurs encourageants, ce qui nous a conduit à réaliser un deuxième prototype immersif.

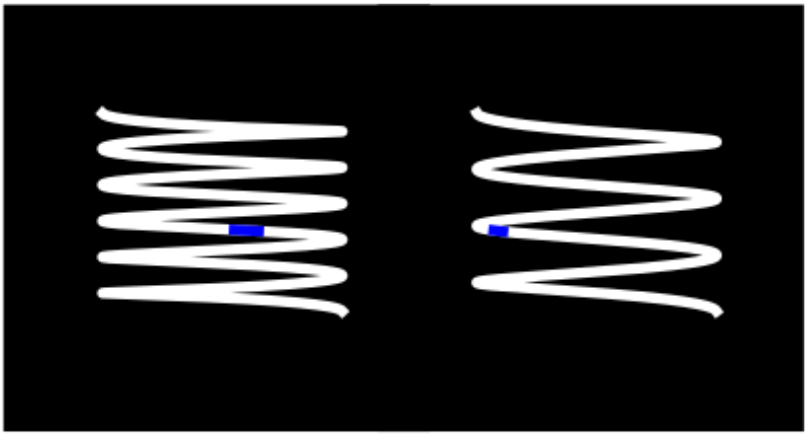


FIGURE 4.2 – La position des points de l'hélice reste constante par rapport à l'axe temporel, quelle que soit la période.

4.3.1 Contexte

Jeu de données choisi et influence sur la conception du prototype

Pour tester l'utilisabilité du prototype sur des scénarios simples, nous avons utilisé le MovieLens 100K Dataset [1]. Cet ensemble de données a été utilisé par Webga et d'autres [129] pour simuler une attaque de fraude sur les évaluations de films. Parce que la détection de fraude et la détection de violation ont de nombreuses similitudes, cet ensemble de données est utilisable pour notre cas d'utilisation. Il se compose de notes données aux films par différents utilisateurs. L'ensemble de données comprend plus de 100 000 votes de 943 utilisateurs sur 1682 films datés sur une période de 7 mois (d'octobre 97 à avril 98). Chaque utilisateur a évalué au moins 20 films et les notes sont des entiers compris entre 1

1. https ://grouplens.org/datasets/movielens/100k/

et 5 inclus. De plus, pour chaque vote, des informations sont fournies sur les utilisateurs : leur âge, leur genre, leur situation professionnelle et leur code postal.

Demande industrielle

La réalisation de prototype a été suivie par deux partenaires industriels de la chaire : Frederic Guihery qui représente la société Amossys, spécialisée dans le conseil en sécurité des technologies de l'information, et Christophe Ponchel, représentant pour la chaire d'Airbus Defence and Space. Leur première demande a été de pouvoir comparer les données actuelles à celles d'une attaque connue. Les attaques auxquelles ils sont soumis ressemblent à des requêtes normales de scrappers (programmes informatiques de collectes de données, utilisés par exemple par Google), à de petites variantes près. La comparaison permettra ainsi de repérer des motifs similaires dans les données, et donc faciliterait la détection d'une attaque. Leur deuxième demande est de pouvoir croiser les informations sur certains utilisateurs. Par exemple, si plusieurs utilisateurs tentent d'accéder à un système au même moment, il peut être utile d'avoir accès à leur adresse IP ainsi qu'à la date de création de leur compte pour décider de leur dangerosité. S'ils viennent tous du même endroit et ont créé leurs comptes au même moment, il est probable qu'ils soient suspects. L'utilisation d'une représentation hélicoïdale permet de répondre à ces deux demandes. Comme présenté dans le concept, les points d'une hélice restent toujours au même niveau par rapport à son axe vertical, quel que soit la période sélectionnée. Donc la comparaison entre deux hélices peut se faire facilement grâce à l'axe temporel qu'elles ont en commun. Pour ce qui est des similitudes de comportements, l'axe commun permet aussi de voir les mêmes comportements périodiques se répéter. De plus, l'environnement 3D permettra d'afficher des informations supplémentaires, par exemple des données catégorielles des utilisateurs.

4.3.2 Implémentation

En raison des restrictions de covid, nous avons utilisé un prototype basé sur le web pour faire passer les expériences sur les ordinateurs des participants. Pour explorer la scène, ils se déplaçaient à l'aide des flèches du clavier et regardaient autour d'eux à l'aide de la souris.

Le Cybercopter utilise une hélice qui peut changer de période pour afficher des motifs périodiques dans une scène 3D. Chaque cube de l'hélice représente un pas de temps et la

couleur du cube indique le nombre d'événements au cours de ce pas de temps (bleu foncé, bleu clair, vert, jaune ou rouge). La période de l'hélice peut être modifiée à l'aide d'un curseur situé en bas à droite de l'interface utilisateur.

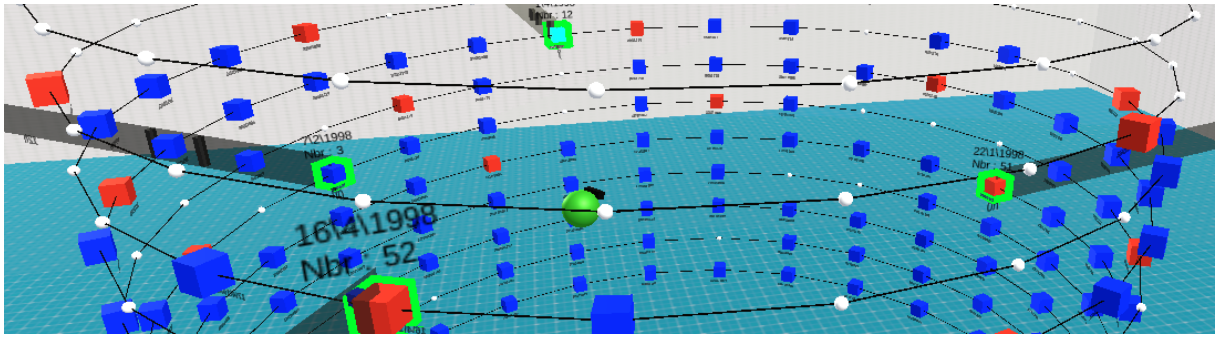


FIGURE 4.3 – Première version du prototype avec un utilisateur à l'intérieur de l'hélice

Lorsque le participant clique sur un cube, des informations supplémentaires apparaissent en suivant le rayon de l'hélice et représentent les distributions d'autres variables au cours de l'intervalle de temps sélectionné. La représentation radiale peut afficher la répartition des votes par heure au cours des journées ou la distribution des notes données au film ce jour. Le participant peut choisir quelles informations afficher en utilisant le bouton « switch » sur l'interface. Comme les représentations radiales sont alignées entre les jours, elles aident à détecter des motifs similaires entre eux (Figure 4.5) en les alignant verticalement. Il ou elle peut également modifier l'angle entre les informations radiales et l'hélice à l'aide d'un curseur sur l'interface utilisateur, ce qui permet de voir des données qui auraient été occultées autrement.

Des informations supplémentaires sont présentes à différents endroits de la représentation. La sélection de cubes, sur l'hélice ou sur le rayon, met également à jour la liste à droite de l'écran. Cette liste montre l'occupation des utilisateurs ayant voté les jours sélectionnés. De plus, les informations sur la date correspondant à un cube sont toujours présentes en dessous de celui-ci, pour avoir des informations supplémentaires sur les journées sélectionnées, l'utilisateur peut passer sa souris par-dessus un cube pour afficher un tooltip contenant le résumé des métriques correspondant à ce jour.

Enfin, l'utilisateur a la possibilité de faire des interactions abstraites avec les données sélectionnées comme modifier le jeu de données en réinitialisant ou modifiant la liste des jours sélectionnés grâces aux boutons de l'interface.

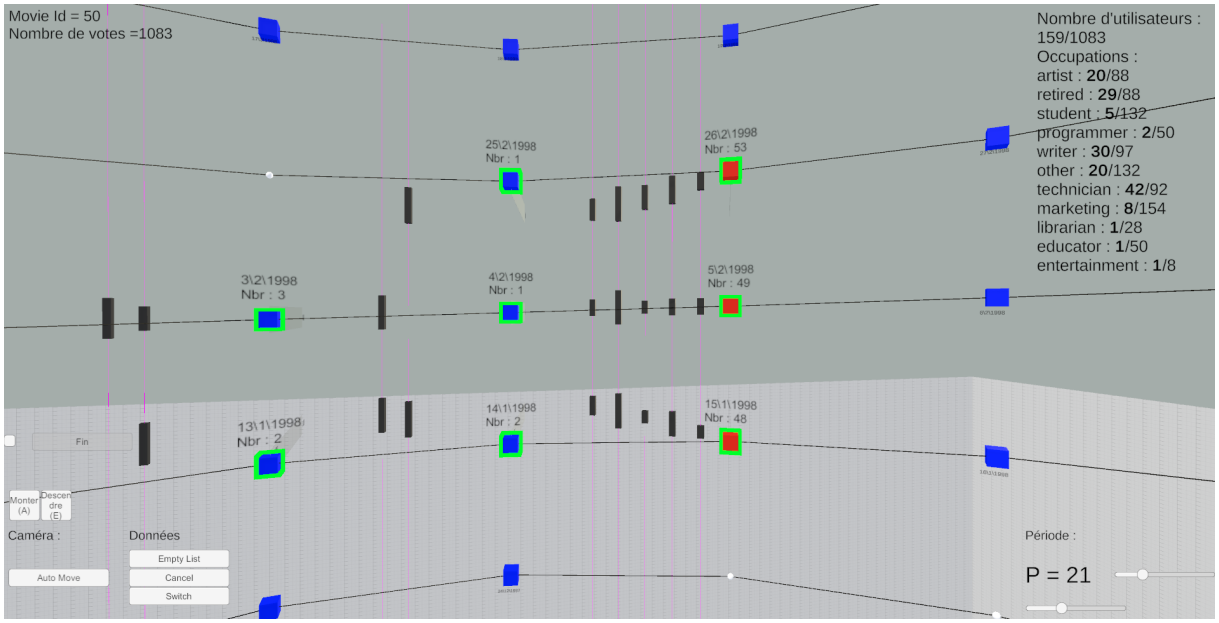


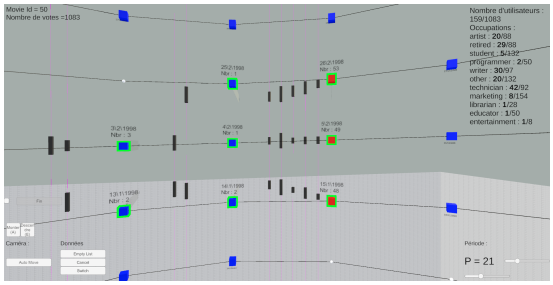
FIGURE 4.4 – Des informations supplémentaires apparaissent à droite de l'écran lorsque l'utilisateur clique sur un cube.

4.3.3 Évaluation

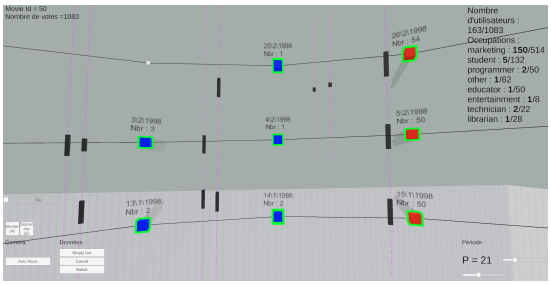
Notre principal objectif est de tester la capacité de nos concepts à être compris et utilisés par des personnes novices à la fois en cybersécurité et en systèmes d'interactions. Ainsi, les participants recrutés n'étaient pas des experts en cybersécurité. Nous avons choisi un scénario simple pour qu'ils puissent le comprendre en peu de temps. Ce scénario a été validé par nos partenaires industriels comme étant représentatif des scénarios de cybersécurité. De plus, nous avons choisi de mesurer l'utilisabilité de notre système à l'aide de la System Usability Scale [130] afin de refléter au mieux la facilité d'utilisation de nos concepts.

En raison de la crise sanitaire de la Covid-19, l'expérience s'est déroulée en téléprésence sur le navigateur web de l'ordinateur du participant. Les participants ont utilisé leur souris et leur clavier pour interagir avec l'environnement 3D. Malgré la variabilité des systèmes, il n'y a pas eu de problème de performance. Après avoir signé un formulaire de consentement en ligne, le participant a découvert l'application dans un tutoriel. Pendant l'expérience, les expérimentateurs ont eu accès aux actions du participant et ont pu l'aider s'il rencontrait un problème qui n'entrait pas dans le cadre de l'expérience (par exemple, un bug).

Les participants ont suivi le scénario suivant : une alerte a été déclenchée sur un film spécifique, afin de savoir s'il s'agissait d'une fausse alerte ou non, ils ont examiné



(a) L'histogramme représente la distribution des votes au cours des heures de la journée. La taille des colonnes représente le nombre de votes par heure



(b) L'histogramme représente la distribution des votes au cours des heures de la journée. La taille des colonnes représente le nombre de votes par heure

FIGURE 4.5 – Les informations radiales apparaissent sous forme d'histogramme lorsque l'utilisateur clique sur un cube. Les alignements des colonnes permettent de détecter des similitudes de comportement entre les jours.

les données du film à l'aide d'un Cybercopter. Les utilisateurs malveillants avaient des caractéristiques et des comportements communs : ils avaient tous le même travail, et ils notaient tous en même temps, avec la même marque qui suggérait la fraude. Le scénario s'est terminé lorsque le participant a jugé s'il y avait ou non fraude.

Après l'expérience, ils ont répondu aux différents questionnaires via un formulaire en ligne. En plus des observations faites pendant l'expérience, le temps de réponse a été mesuré ainsi que la facilité d'utilisation du prototype à l'aide du questionnaire SUS.

4.3.4 Résultats et leçons retenues

Huit participants âgés de 23 à 30 ans, 5 femmes et 3 hommes, ont pris part à l'expérience. Sept d'entre eux ont tiré la sonnette d'alarme en moins de 10 minutes (moyenne=11min27sec, médiane=8min06sec, écart type=12min40sec, max=42min27sec, min=4min23sec), celui qui a mis le plus de temps était le seul à ne pas avoir d'expérience dans le domaine des jeux vidéos. Bien que le nombre de participants soit faible, le score SUS moyen de 77 (moyenne=77, médiane=79, écart type=11, max=90, min=55), est un indicateur encourageant de la facilité d'utilisation du Cybercopter.

Nous avons choisi de faire apparaître les utilisateurs au centre de l'hélice, ce qui les a perturbés pour comprendre la représentation. En général, leur première action consistait à quitter l'hélice afin d'observer l'environnement depuis l'extérieur, ce qui leur permettait d'obtenir une meilleure perception de sa forme générale. Par la suite, ils ajustaient la

période pour identifier les motifs périodiques, puis sélectionnaient des cubes d'intérêt avant de se déplacer pour se rapprocher des motifs présents dans les représentations radiales. Par conséquent, nous estimons qu'il est nécessaire de concevoir une hélice plus petite et plus facile à manipuler, décentrée par rapport à l'utilisateur. De plus, l'ajout de plusieurs hélices pour la comparaison renforce l'intérêt du point de vue excentrique. Certaines personnes n'étaient pas familières avec la navigation dans un environnement en trois dimensions, ce qui entravait leur mobilité et, par conséquent, leur expérience globale. Cela nous a conduits à opter pour un prototype immersif offrant des interactions intuitives et des représentations hélicoïdales exocentrées.

4.4 Prototype II

À la suite de l'analyse des retours d'expérience recueillis grâce au premier prototype, nous avons pris la décision de concevoir un deuxième prototype immersif, fondé sur des multiples de représentations hélicoïdales pour la visualisation de données périodiques. Cette section présente le cas d'utilisation que nous avons sélectionné pour évaluer ce prototype, puis elle détaille l'implémentation de ce dernier ainsi que de l'interface 2D employée pour les comparaisons. Nous exposons également notre protocole expérimental dans lequel des utilisateurs ont été chargés de trier des alertes émises par des capteurs. Enfin, nous exposons les résultats obtenus ainsi que les enseignements que nous avons tirés de cette expérience.

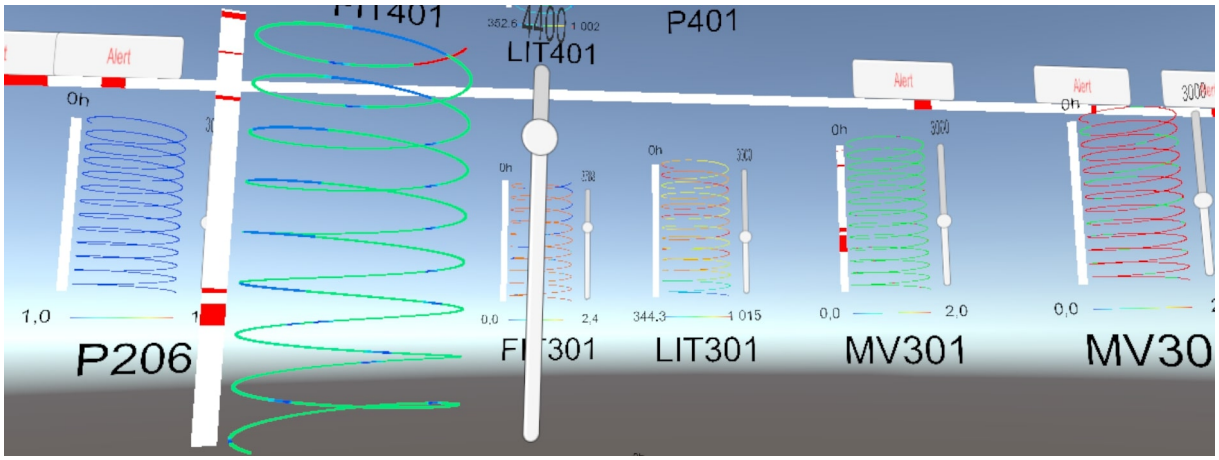


FIGURE 4.6 – Vue d'ensemble du deuxième prototype

4.4.1 Contexte

L'un des défis de l'industrie 4.0 consiste à détecter les attaques sur les réseaux de technologie opérationnelle (OT) après que les attaquants se déplacent latéralement à partir d'un réseau de technologie de l'information (IT) compromis. Les réseaux IT et OT étaient auparavant séparés, mais avec l'émergence de nouvelles technologies telles que l'Internet des objets, les réseaux OT ont tendance à être connectés à Internet, ce qui peut les rendre vulnérables aux attaques. Cependant, il existe peu de solutions pour détecter les attaques sur les réseaux OT, une façon de le faire est d'utiliser les données déjà disponibles, telles que celles provenant de capteurs physiques. En effet, un comportement suspect des captuers physiques peut indiquer une attaque (bien qu'il puisse également être dû à un dysfonctionnement mécanique). Pour détecter ces valeurs suspectes, Lofhink et al. ([121]) décrivent un système de visualisation basé sur des spirales pour trier les alarmes levées par des profils matriciels sur les capteurs d'une station d'épuration. Nous avons choisi ce cas d'utilisation pour évaluer l'utilité d'un essaim de Cybercopters, car les capteurs affichent des signaux périodiques. Afin de donner un aperçu de ce cas d'utilisation, nous décrivons dans cette section l'ensemble des données utilisé, le fonctionnement de l'algorithme de détection des anomalies et les exigences du système de tri visuel.

Secure Water Treatment (SWaT) dataset

Le jeu de données SWaT [2] ([131]) décrit les données (IT et OT) d'un banc d'essai d'une station d'épuration sur une période de onze jours. La station d'épuration fonctionne normalement pendant les sept premiers jours de l'ensemble de données. Le huitième jour, des cyberattaques sont lancées contre la station et peuvent perturber son fonctionnement. Nous nous intéressons aux données provenant des capteurs et des actionneurs de la partie physique de l'ensemble de données (tableau 4.1). Elles couvrent les 6 processus de l'usine, de l'arrivée de l'eau à sa distribution dans le circuit. Les 51 capteurs et actionneurs présentent un comportement périodique qui peut être modifié lors d'attaques visant à perturber le fonctionnement nominal de l'usine. Pour cette expérience, nous avons pris 10000 secondes (2 h 50 min) du fonctionnement nominal de l'usine et 27700 points (7 h 40 min) du fonctionnement sous attaques (figure 4.7).

Timestamp	Sensors				
	DPIT301		MV301		...
	Value	Score	Value	Score	...
1	143	0.1	1	0	...
2	254	0.1	1	0	...
...	...	...	...	...	...
3154	27	0.9	2	0.95	...
3155	28	0.85	3	0.93	...
...	...	...	...	...	...
...	...	...	...	...	...

TABLE 4.1 – Nous appliquons l'algorithme de matrix profile pour détecter les anomalies sur chaque capteur de l'ensemble de données SWaT. Chaque capteur a une valeur spécifique et un score d'anomalie associé à chaque horodatage (une par seconde)."

Matrix Profiles

Les matrix profiles sont des algorithmes de découverte de motifs et de détection d'anomalies. Ils ne nécessitent pas d'entraînement avant d'être utilisés sur un ensemble de données et ont peu de paramètres à régler ([132]). Ces algorithmes créent des fenêtres temporelles pour chaque pas de temps de l'ensemble de données. Chaque fenêtre est comparée aux autres, ce qui donne un score de similarité compris entre 0 (le motif présent dans cette fenêtre peut être trouvé ailleurs) et 1 (le motif présent dans cette fenêtre est le seul qui existe dans l'ensemble de données). Plus précisément, l'algorithme calcule un matrix profil associé à une série temporelle, créant ainsi un vecteur qui enregistre les distances euclidiennes entre chaque sous-séquence de la série temporelle et sa sous-séquence voisine la plus proche.

Algorithm 1 Matrix profile

Entrée : Une série temporelle T, longueur de sous-séquence d'intérêt m
Sortie : Un matrix profile P et son index associé I

```
n ← Length(T)
P ← infs, I ← zeros, idxes ← 1 :n-m+1
for each idx in idxes do
    D ← MASS(B[idx], TA) [3]
    P, I ← ElementWiseMin(P, I D, idx)
end for
return  P, I
```

2. https ://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/

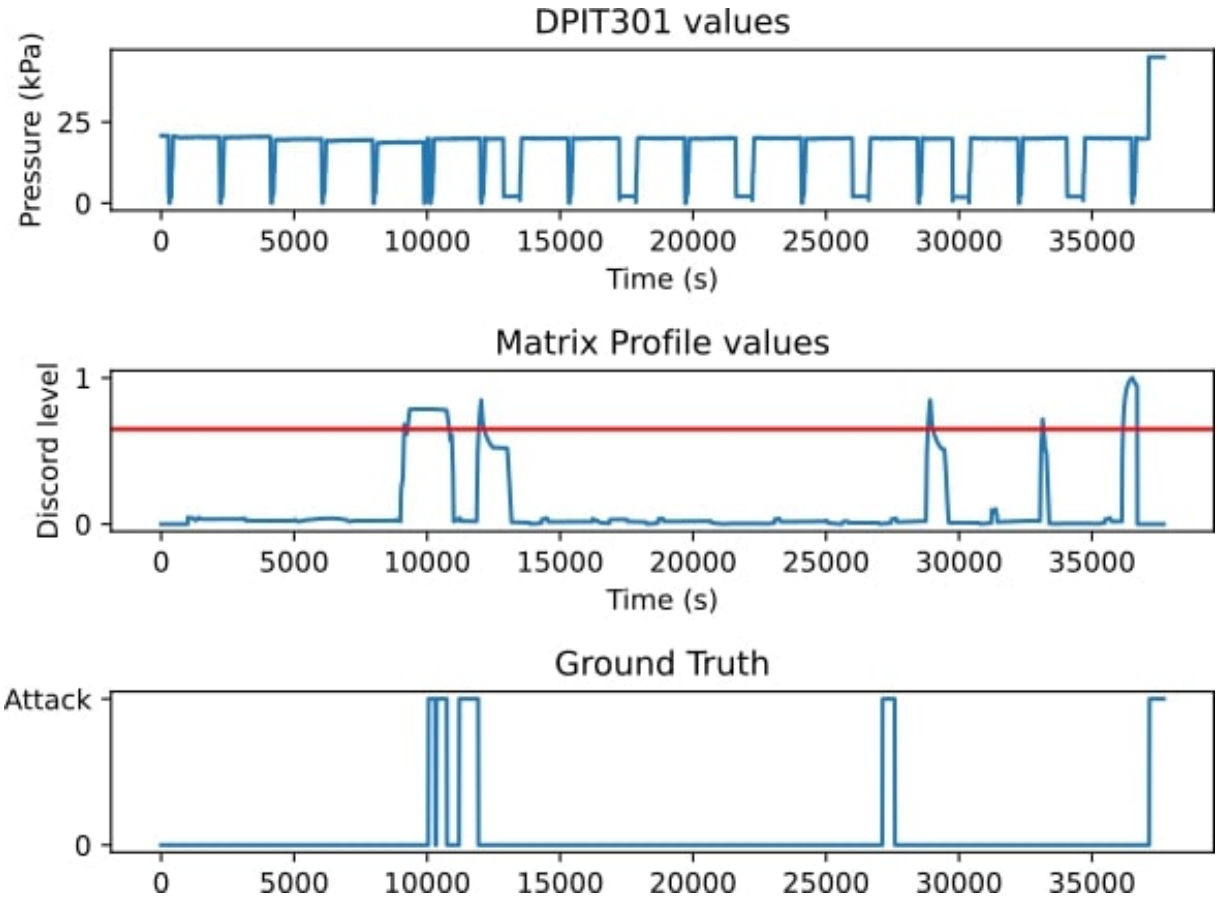


FIGURE 4.7 – L'algorithme de matrix profile appliqué au capteur DPIT 301 et à la vérité terrain des attaques ([131])

Plus le score d'un événement est proche de 1, plus il est probable que cet événement soit une anomalie ([133]). Lohfink et Al. ([121]) définissent trois catégories de score : le type I où une anomalie est improbable, le type II où le score est plus élevé que la normale, mais pas assez pour être considéré comme une anomalie certaine, et le type III où le score est suffisamment élevé pour être une anomalie certaine. Les seuils entre ces catégories sont fixés en utilisant la partie de l'ensemble de données sans attaque afin de générer le moins de faux positifs possible. Cependant, cette méthode de détection des anomalies est vouée à générer des fausses alertes, ce qui rend nécessaire un outil de visualisation pour les catégoriser.

Besoins

Pour évaluer et comparer nos solutions, nous utilisons cinq des six exigences exprimées par Lohfink et al. ([121]) à la suite d'entretiens avec des experts.

R1 : La surveillance du système et l'analyse du triage doivent être prises en charge simultanément.

R2 : Les anomalies détectées doivent être clairement mises en évidence dans les données.

R3 : Le système de visualisation doit permettre d'identifier les faux positifs.

R4 : Le système de visualisation doit permettre d'identifier les faux négatifs.

R5 : Le système de visualisation doit permettre aux experts en cybersécurité et aux non-spécialistes d'effectuer des analyses de triage.

Nous avons supprimé l'exigence : *Classification des valeurs de la catégorie II comme anormales ou normales*, car nous n'utilisons pas les alertes de la catégorie II dans cette expérience. En effet, nous avons regroupé la catégorisation des alertes de la catégorie II et de la catégorie III en une seule tâche : catégoriser un comportement anormal détecté comme un vrai ou un faux positif. Comme l'indiquent les auteurs, la catégorie II peut être omise si un intervalle "tampon" n'est pas nécessaire.

4.4.2 Implémentation

Sur la base du Cybercopter, nous avons créé l'essaim Cybercopters (Figure 4.6) qui met en œuvre de multiples représentations hélicoïdales pour mettre en évidence les comportements périodiques dans les données temporelles.

La première différence entre les deux prototypes est l'utilisation d'un casque de réalité virtuelle pour notre deuxième prototype. La parallaxe permise grâce au suivi des mouvements de la tête du casque aide à mieux comprendre les visualisations en 3D. De plus, ce prototype affiche réellement les alertes levées sur un ensemble de données provenant de la communauté de la cybersécurité.

De plus, dans le cas de l'essaim de Cybercopters, nous avons utilisé le fait que d'autres représentations peuvent être liées spatialement à une hélice pour représenter la chronologie des alertes levées sur chaque capteur à côté d'eux (figure 4.8). Lorsqu'une alerte est déclenchée sur un capteur, la barre d'alerte devient rouge à l'heure correspondante. Ainsi, l'utilisateur peut établir un lien entre une rupture suspecte dans les schémas et l'alerte à la même hauteur. Cette représentation garantit que nous n'utilisons pas la même dimension

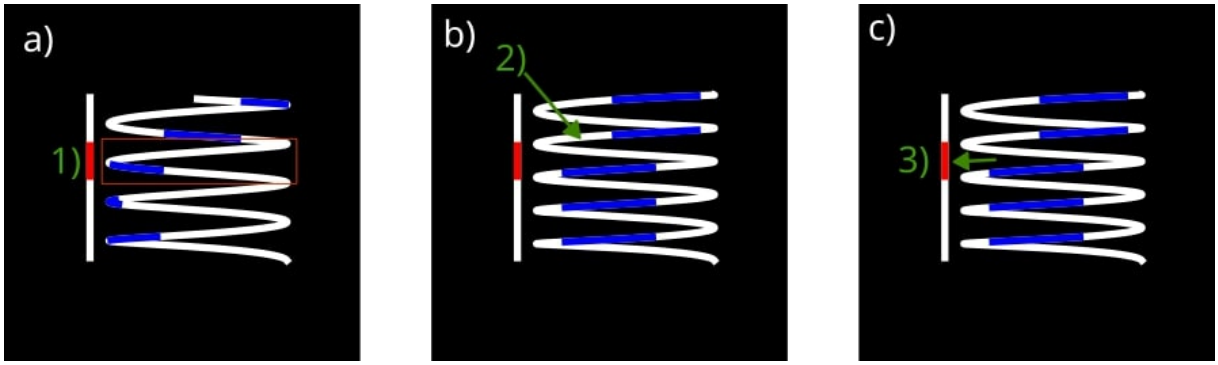


FIGURE 4.8 – L'hélice permet de laisser un espace pour représenter les alertes sur une représentation adjacente. Ainsi, la représentation de la zone anormale n'interfère pas avec la représentation des données du capteur. Cela facilite le travail des opérateurs qui peuvent : a) détecter la zone des activités suspectes, b) se concentrer sur l'hélice pour faire apparaître une rupture dans un schéma périodique, c) contrôler si la rupture correspond à l'alarme.

visuelle pour représenter deux dimensions différentes des données : les valeurs des capteurs et les valeurs des alertes. Ainsi, l'utilisateur a toujours accès aux deux informations et l'épaisseur de la ligne peut être utilisée pour représenter une autre dimension de l'ensemble de données. De plus, comme deux hélices ayant différentes périodes auront toujours leurs points de données à la même position verticale, cela peut simplifier la comparaison des comportements de deux capteurs.

Nous avons utilisé l'essaim de Cybercopters pour visualiser les données SWaT, chaque capteur et actionneur a des schémas périodiques et peut être associé à des alertes. Dans ce but, chaque capteur est représenté par un Cybercopter avec sa propre période. Nous avons choisi de représenter 10h30min du jeu de données en nous basant sur ([121]), plus de points de données pourraient être affichés si nécessaire. La période de l'hélice peut être réglée par l'utilisateur à l'aide du curseur situé sur le côté de l'hélice, ou en pointant l'hélice avec sa main tout en utilisant le joystick d'un contrôleur, pour lui permettre d'être plus précis sans avoir à bouger les bras. En déplaçant une main dans l'hélice, l'utilisateur peut obtenir l'horodatage de ce point pour déterminer l'heure d'un événement. Chaque capteur possède sa propre légende qui dépend de ces valeurs sur l'ensemble de la plage de données. L'utilisateur peut faire pivoter tous les points de données le long de l'axe vertical de l'hélice s'il le souhaite afin de mieux aligner une alerte sur un comportement suspect dans les données. Cette interaction permet également de modifier la vue de l'hélice si certaines parties sont occultées par les flèches à l'avant [134].

Afin de donner une vue d'ensemble de la situation, une grande barre d'alerte regroupant toutes les alertes est présente à l'horizon (Figure 4.6). Lorsque l'utilisateur souhaite savoir quel(s) capteur(s) correspond(ent) à quelle alerte, la représentation utilise un signal audio couplé à un signal visuel pour indiquer à l'utilisateur les capteurs correspondant à une alerte. Lorsque l'utilisateur sélectionne une alerte, un son géo-spatialisé provenant du capteur correspondant lui permet de localiser sa position approximative, puis il peut localiser le capteur plus précisément grâce au changement de couleur du nom du capteur. Si plusieurs capteurs déclenchent une alarme, leurs sons sont joués l'un après l'autre.

Les hélices peuvent être présentées comme un "mur", mais d'autres dispositions peuvent être utilisées, comme un demi-cercle si le nombre de capteurs augmente ([111]). Le placement des hélices dans l'environnement 3D peut également se faire en fonction de la position du capteur dans l'environnement réel ([126]), ou en fonction de la place du capteur dans le processus d'une manière plus abstraite. La métaphore de l'hélice peut être étendue à la réalité augmentée, par exemple pour transmettre l'état de chaque capteur à l'intérieur de la station d'épuration aux opérateurs de la station (**ElSayed2015a**).

Nous avons choisi de placer les participants en position assise, car cela correspond à la position d'un analyste en cybersécurité travaillant à un bureau. De plus, selon Wagner et Al. [135], la position assise n'affecte pas négativement les performances dans l'IA, mais elle peut augmenter les performances physiques et mentales, tout en augmentant l'effort requis pour accomplir la tâche. C'est également plus pertinent dans la comparaison avec l'interface 2D où les participants sont également assis. Selon Krause et Al. ([136]), une table de RV, c'est-à-dire un endroit où les visualisations sont représentées et avec lesquelles on interagit, réduit considérablement le nombre de déplacements nécessaires pour accomplir une tâche. Toutefois, selon Liu et Al. [111], il semble que les dispositions en arc de cercle soient meilleures pour représenter un grand nombre de représentations. Nous avons choisi un compromis entre les deux, où toutes les hélices sont représentées dans une disposition radiale et où les participants peuvent sélectionner les hélices intéressantes avec lesquelles interagir devant eux. Ce choix crée le besoin d'un outil de sélection pour les objets distants, d'où l'utilisation du laser pour amener les capteurs au participant. L'utilisateur peut sélectionner une hélice en la pointant avec le rayon virtuel à l'extrémité de sa main et l'hélice entre directement dans la main en cliquant sur un bouton. Grâce à cette interaction, il peut rester assis pendant tout le processus de visualisation et n'a pas besoin de se déplacer dans l'environnement 3D. Il peut organiser son environnement à sa convenance en fonction de la tâche qu'il souhaite effectuer : zoomer sur le capteur

qui l'intéresse, regrouper plusieurs capteurs ayant un comportement similaire ou jeter un capteur devenu inutile derrière lui, par exemple. Enfin, l'utilisateur peut modifier la taille de l'hélice en la saisissant à deux mains et en les écartant. En ce qui concerne les interactions, nous avons choisi d'utiliser les contrôleurs de base du casque de Réalité Virtuelle, car c'est une pratique courante dans la communauté de l'Immersive Analytics [89].

Nous avons utilisé le fait que d'autres représentations peuvent être liées spatialement à une hélice pour représenter la chronologie des alertes levées sur chaque capteur à côté d'eux (figure 4.1). Lorsqu'une alerte est déclenchée sur un capteur, la barre d'alerte devient rouge à l'heure correspondante. Ainsi, l'utilisateur peut établir un lien entre une rupture suspecte dans les schémas et l'alerte à la même hauteur. Cette représentation garantit que nous n'utilisons pas la même dimension visuelle pour représenter deux dimensions différentes des données : les valeurs des capteurs et les valeurs des alertes. Ainsi, l'utilisateur a toujours accès aux deux informations et l'épaisseur de la ligne peut être utilisée pour représenter une autre dimension de l'ensemble de données.

Pour développer Cybercopter, nous avons mené deux expériences pilotes au cours desquelles les participants nous ont aidé à améliorer la convivialité et l'interface de notre prototype. La taille des hélices et leur distance par rapport à l'utilisateur ont été privilégiées par ces premiers participants, qui ont également suggéré d'ajouter une fonction de rotation à l'hélice sans faire pivoter la barre d'alerte. Un des premiers testeurs a suggéré d'utiliser les joysticks pour interagir avec l'hélice, à la fois pour changer la période et pour faire tourner l'hélice, car il pensait que c'était la solution la plus intuitive. Cependant, nous ne voulions pas trop d'interactions complexes pour ne pas être trop biaisés par rapport à la 2D. Mais il pourrait être intéressant de mener une seconde expérience sur les interactions avec ce type de représentations.

Implémentation et passage à l'échelle

L'essaim de Cybercopter et le prototype 2D ont été développés avec unity 2019.3.7f1 dans C#. La version VR utilise Steam VR [4] et tilia [5] pour des interactions spécifiques. Le traitement des données est réalisé avec python 3.7 en utilisant la bibliothèque matrix profile développée par la Matrix Profile Foundation ([137]). Les données sont transférées de python à unity via des fichiers .json. Bien qu'elle soit réalisée de manière asynchrone

4. https ://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647
5. https ://www.vrtk.io/tilia.html

pour l'instant, cette implémentation peut être utilisée pour traiter les données en continu dans le cadre de travaux futurs.

4.4.3 Évaluation

Pour évaluer l'utilité de Cybercopters, nous le comparons à une interface 2D basée sur les travaux de Lohfink et al. ([121]). Ils ont utilisé une interface basée sur une spirale pour afficher des motifs périodiques dans les données des capteurs du SWaT.

Prototype 2D

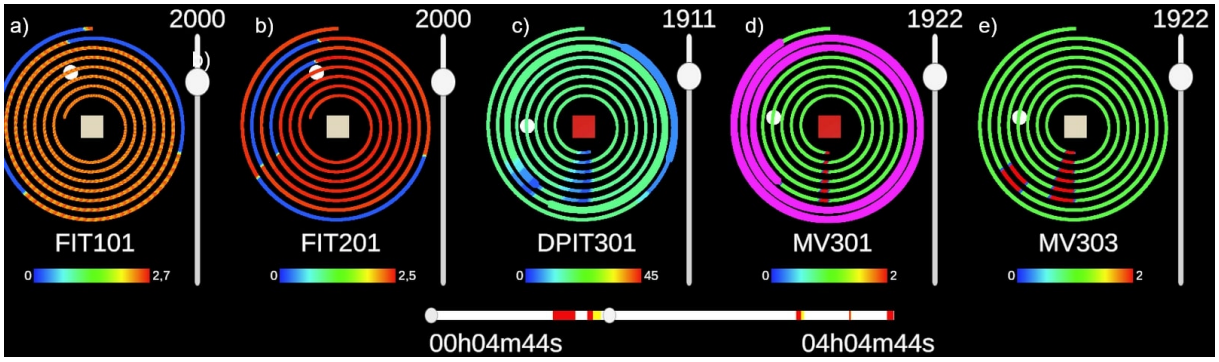


FIGURE 4.9 – Les cinq capteurs de *set1* représentés par l'interface 2D. Au milieu, la ligne du temps, qui permet aux utilisateurs de sélectionner la période affichée. En bas, l'interface pour répondre aux alertes. Nous pouvons voir un faux négatif sur MV303 car il y a une rupture claire dans le motif périodique. Un clic sur le bouton rouge au centre de la spirale de DPIT301 met en évidence les deux alertes présentes dans la période sélectionnée. Lorsqu'elles ne sont pas mises en évidence, l'épaisseur des spirales indique la présence d'alertes.

L'interface 2D est divisée en deux parties : la timeline qui affiche les alertes sur l'ensemble des données et permet de sélectionner la fenêtre de temps affichée sur les graphiques en spirale (figure 4.9). Les alertes affichées sur la ligne de temps représentent les alertes de tous les capteurs. Inspirée par ([121]), la fenêtre temporelle couvre un maximum de 14 400 valeurs, ce qui équivaut à 4 heures de logs. Ce maximum peut être modifié si nécessaire. Pour naviguer dans l'ensemble de données, les utilisateurs peuvent faire glisser l'un ou l'autre côté du curseur ou simplement cliquer sur la zone qui les intéresse pour déplacer la fenêtre temporelle.

La période de chaque spirale peut être modifiée séparément à l'aide du curseur situé à droite et de la molette de la souris pour des changements plus précis. Pour simplifier

la compréhension de l'interface par les participants, la couleur des points de données suit une carte thermique en fonction de la valeur du capteur sur l'ensemble des données et pas seulement sur la fenêtre. Les légendes sont toujours visibles et affichent les valeurs minimales et maximales de chaque capteur. Pour l'expérience, nous n'avons représenté que cinq capteurs à la fois, mais 10 peuvent être affichés en même temps sur la même fenêtre sans poser de problèmes de performance.

Pour afficher les alertes levées sur un capteur, l'épaisseur de la spirale est augmentée au moment de l'alerte, ce qui attire l'œil sur les zones critiques. De plus, le carré d'alerte au centre de la spirale devient rouge si les données du capteur contiennent une alerte sur la plage de temps sélectionnée. Pour mettre en évidence les alertes, le participant peut cliquer dessus pour les faire clignoter en magenta (figure 4.9).

Lorsque la souris passe sur une spirale, un point lumineux apparaît au-dessus de celle-ci et sur le pas de temps correspondant dans les autres spirales pour faciliter la comparaison, car le même pas de temps ne se trouve pas au même endroit dans deux spirales ayant des périodes différentes.

Le prototype 2D a également fait l'objet de projets pilotes qui ont permis d'améliorer son interface, notamment en déterminant la bonne taille des spirales et des curseurs.

Apparatus

L'évaluation des deux conditions a été réalisée sur le même appareil : un ordinateur portable doté d'un écran de 15 pouces avec une résolution de 1920x1080 pixels. L'ordinateur portable dispose de 32 Go de RAM, d'un processeur *Intel Core i7*, d'une carte graphique *Nvidia Quadro RTX 3000* et est équipé du système d'exploitation *Windows 10*. La version 2D est utilisée en plein écran et peut afficher 10 capteurs de manière fluide, les participants interagissant avec elle à l'aide de la souris. La version VR utilise le Vive Cosmos et ses contrôleurs et peut afficher les 51 capteurs dans la scène et interagir avec eux sans ralentissement. Dans les deux cas, les participants sont assis devant l'ordinateur.

Participants

Nous avons recruté 24 volontaires non rémunérés (3 femmes, 21 hommes) âgés de 23 à 50 ans (moyenne : 29,5, écart-type : 7,4), dont 75 % avaient déjà une expérience de l'environnement 3D (jeux vidéo, CAO, modélisation 3D...). Nous leur avons ensuite demandé d'évaluer sur une échelle de 1 (novice absolu) à 5 (expert) leur habitude de la réalité virtuelle (moyenne : 2,45, sd : 1,35) et de la cybersécurité (moyenne : 2, sd : 1,45).

Tous les participants ont effectué au moins cinq années d'études dans un domaine scientifique. Même si seuls quelques-uns d'entre eux avaient une expertise en cybersécurité, la tâche était suffisamment compréhensible pour que des personnes ayant une solide formation scientifique puissent la comprendre. Trois experts en cybersécurité ont été interrogés séparément pour donner leur avis, car ils n'étaient pas assez nombreux pour que les résultats soient statistiquement significatifs. Un participant daltonien n'a eu aucun problème à répondre à l'expérience.

Collecte des données

Au début de l'expérience, chaque participant répond à un questionnaire démographique. Dans cette étude, la variable indépendante est la méthode de représentation des données. Ce facteur a deux valeurs possibles dans ce contexte : 2D avec spirales (*2DSpirales*) et 3D avec hélice (*3DHélice*). Les variables dépendantes sont divisées en deux catégories : les mesures quantitatives objectives (temps de réponse aux alertes et pourcentage de bonnes réponses) et les mesures qualitatives subjectives provenant de questionnaires (System Usability Scale ([130]), Raw NASA TLX ([138]), et SHORT Flow State Scale ([139])). Les détails de ces questionnaires sont disponibles en annexes.

Protocole

Présentation : Le participant reçoit une explication du contexte de l'expérience, des données utilisées et de leurs caractéristiques périodiques. Afin de comprendre pourquoi il y a des fausses alertes, le participant reçoit une brève description du fonctionnement de l'algorithme du profil matriciel et de ses limites. Le participant apprend que son objectif au cours de l'expérience est de traiter les alertes et de distinguer les vrais des faux positifs. Le participant est libre de poser toutes les questions qu'il juge nécessaires pour dissiper ses doutes. Il signe ensuite le formulaire de consentement éclairé.

Tutoriel : Le participant effectue une séance de formation afin de se familiariser avec la première interface utilisée. Pour les deux interfaces, les mêmes données sont utilisées. Le participant est d'abord invité à trouver un motif périodique dans un ensemble de données simple. Il examine ensuite quatre alertes : une vraie positive due à une rupture du schéma périodique, une due à une surtension et deux fausses positives. Au cours de ces tâches, toutes les fonctionnalités de l'interface lui sont présentées, et il s'entraîne en utilisant les boutons de réponse.

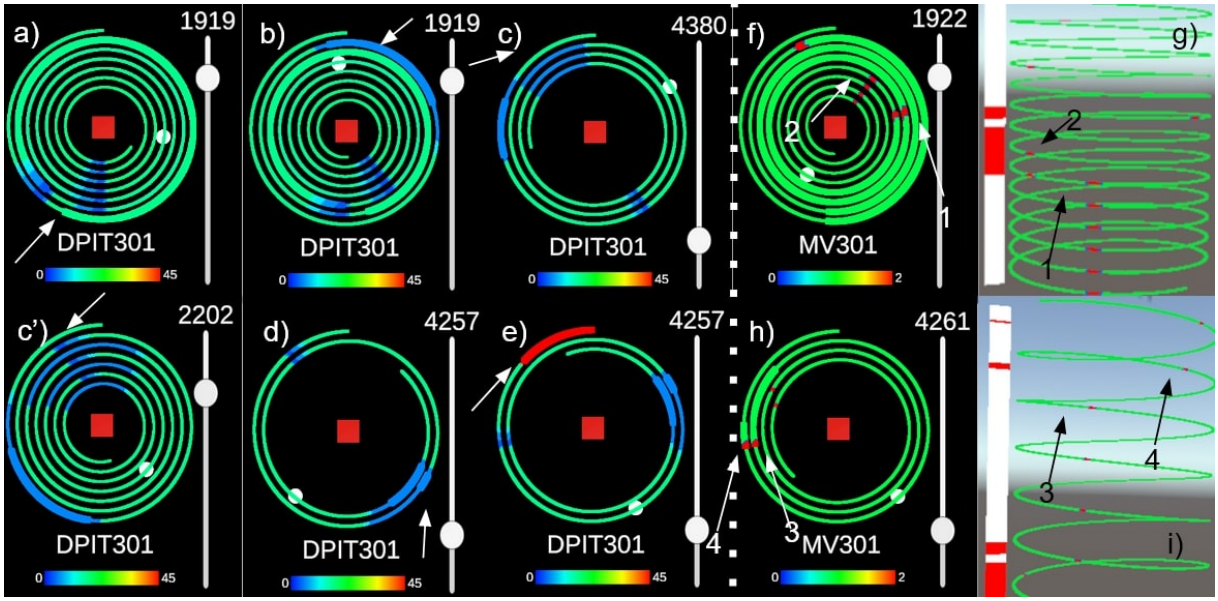


FIGURE 4.10 – Les alertes de *sets1* sont représentées par l'interface 2D (a, b, c, c', d, e). c') représente la même alerte que c) avec une période réduite de moitié. Le schéma "un long, un court" (p = 2200s) de c') peut être détecté aussi facilement que le schéma "deux pics" (p = 4400s) de c). Les alertes de *set 2* sont représentées avec les deux interfaces, f) et g) représentent les alertes 1 et 2, g) et i) représentent les alertes 3 et 4..

Tâches : Le participant doit effectuer une tâche de traitement des alertes avec ce prototype. Pour l'expérience, nous ne représentons que 5 capteurs, sans la barre d'alerte générale, et nous leur demandons de juger les alertes d'un capteur. Comme il y a deux séries d'alertes, la moitié des participants commence par *ensemble 1*, et l'autre moitié par *ensemble 2*. *ensemble 1* est composé de FIT101, FIT201, DPIT301, MV301, MV303 et *ensemble 2* de LIT101, P101, AIT202, P205, MV301 (figure 4.10). Les ensembles sont choisis de manière que les capteurs présentent des schémas similaires bien que des valeurs différentes, car ils font partie du même processus. Pour la détection des ruptures de périodicité, la fenêtre de l'algorithme de profil matriciel est fixée à 2000 pour DPIT301 et à 4400 pour MV301, le seuil de déclenchement des alarmes est fixé à 0,65.

Afin de tenir compte de l'effet d'apprentissage sur la tâche en question, nous avons combiné l'ordre des prototypes 2D et 3D et l'ordre de *set 1* et *set 2*. Les participants sont donc divisés en quatre groupes de six personnes qui ont effectué les tâches comme suit :

G1 : premièrement le *set 1* avec le prototype 2D, puis le *set 2* avec le prototype 3D.
G2 : premièrement le *set 2* avec le prototype 2D, puis le *set 1* avec le prototype 3D.
G3 : premièrement le *set 1* avec le prototype 3D, puis le *set 2* avec le prototype 2D.

G4 : premièrement le *set 2* avec le prototype 3D, puis le *set 1* avec le prototype 2D.

Premièrement, les participants ont dû détecter un faux négatif dans MV303 pour *set 1* et AIT202 pour *set 2*, c'est-à-dire détecter une rupture de motif périodique dans le comportement d'un capteur pour lequel aucune alerte n'a été déclenchée. Avant de donner le nom du capteur, ils ont appuyé sur le bouton *start* de *Motif 1*, puis sur le bouton *stop* lorsqu'ils ont donné une explication de leur réponse, ce qui permet de mesurer leur temps de réponse. Cette partie évalue *R4* : la capacité à détecter les faux négatifs dans le système. Ensuite, il leur a été demandé de détecter un faux négatif dans FIT 201 pour *set 1* et dans P101 pour *set 2* dans une partie sans motif particulier. En effet, un premier pilote que nous avions mené nous avait montré que les participants sont réticents à déclarer qu'ils n'ont rien trouvé, et nous voulions les préparer aux faux positifs dans les alarmes. Les alertes de ces quatre capteurs avaient été supprimées pour créer des faux négatifs.

Enfin, les participants ont répondu aux alertes levées sur un seul capteur, car nous voulions évaluer l'efficacité de la visualisation 3D par rapport à la visualisation 2D plutôt que l'efficacité d'un essaim. Nous nous sommes appuyés sur les avis d'experts concernant l'essaim. Le prétraitement des alarmes est effectué avant l'expérience à l'aide du même algorithme de matrix profil. Elles se produisent à des moments différents sur le même capteur, mais sont toutes visibles sur le capteur au début de l'expérience. Les participants pouvaient choisir de répondre aux alertes dans l'ordre de leur choix, mais ils ont tous répondu dans l'ordre chronologique. Ils ont été informés que toutes les alertes se produisaient sur le même capteur pour chaque ensemble, DPIT301 pour *ensemble 1* et pour MV301 *ensemble 2*. Pour *set 1*, les participants devaient répondre à 5 alertes et pour *set 2*, ils devaient répondre à 4 alertes. Les quatre premières alertes présentaient des caractéristiques similaires et ont été utilisées pour comparer le temps d'exécution. Il n'a pas été possible de trouver la cinquième alerte pour *set 1* dans d'autres capteurs de l'ensemble de données. Toutefois, nous l'avons conservée pour vérifier que les participants étaient capables de détecter une alerte provoquant une augmentation plutôt qu'une interruption des schémas périodiques (figure 4.10). Les alertes *Set 1* ont toutes été déclenchées sur DPIT301 et les alertes *Set 2* ont toutes été déclenchées sur MV301 (figure 4.10). Les participants ont répondu aux alertes en appuyant sur le bouton *vrai positif* ou *faux positif* de l'alerte correspondante. Le chronomètre a démarré lorsqu'ils ont appuyé sur le bouton *start* en haut des 5 alertes. Il a été demandé aux participants de répondre aussi vite que possible tout en fournissant les explications nécessaires à leur résultat afin de simuler une situation dans laquelle un opérateur SOC doit répondre rapidement, mais correcte-

ment à des alertes. Par conséquent, le temps enregistré correspond au temps passé entre la dernière alerte traitée et la conclusion du raisonnement du participant (vrai ou faux positif). Nous pensons que ce temps mesuré est pertinent, car il contient la détection de l'alerte à traiter, la recherche d'une rupture potentielle dans un schéma périodique, ce qui correspond au travail réel d'un opérateur. De plus, il leur a été demandé de répondre complètement à chaque alerte avant d'essayer de répondre à la suivante. Ces questions permettent d'évaluer *R3* : "L'identification des faux positifs devrait être possible grâce au système de visualisation" et *R2* : "Les anomalies détectées doivent être clairement mises en évidence dans les données". Après avoir trié les alertes, les participants ont été interrogés sur les caractéristiques des alertes qu'ils avaient traitées afin d'évaluer *R1* :

— S'ils ont traité la première alerte comme un vrai positif : À quel moment la tendance s'est-elle interrompue ? Quelle était sa période ?

— S'ils ont traité la troisième alerte sur DPIT301 comme un vrai positif : Quelle a été la durée de la tendance avant qu'elle ne se brise ?

Ces questions sont inspirées de celles de Lohfink et al. ([121]) pour évaluer leurs interfaces. Par rapport aux originaux, nous n'avons gardé que les questions sur la quantification des caractéristiques d'une alerte, car les autres questions évaluaient les capacités à détecter les alertes et à les classer comme fausses ou vraies positives, ce qui est déjà couvert dans notre expérience. Après les tâches, chaque participant a répondu aux questionnaires et une phase de discussion/question ouverte a eu lieu. Il a répété le tutoriel, les tâches et les questionnaires avec l'autre *set* et l'interface. À la fin de l'expérience, on lui a également demandé si l'un des ensembles était plus difficile que l'autre, et si oui, lequel.

Hypothèses

Nous avons formulé les hypothèses suivantes concernant les deux représentations :

H1 : Les deux conditions sont efficaces pour la détection des ruptures dans les motifs périodiques, ce qui est mesuré par le taux de réussite et le temps de réponse. Comme les deux visualisations sont destinées à afficher des signaux périodiques, elles sont toutes deux censées réussir à accomplir cette tâche.

H2 : Les deux conditions ont le même taux de réussite pour les réponses aux alertes. Les deux visualisations étaient destinées à corréler les alertes avec les ruptures dans les signaux périodiques. Les participants devraient être aussi efficaces l'un que l'autre avec les deux interfaces.

H3 : *3DHelix* est plus rapide que *2DSpirals* pour les réponses aux alertes. Le prototype

3D étant destiné à faciliter les corrélations entre les alertes et les signaux, il devrait améliorer le temps de réponse des participants par rapport au prototype 2D.

H4 : *2DSpirals* demande moins d'efforts que *3DHelix*, ce qui est mesuré par les réponses au NASA-TLX brut. Étant donné que les interfaces 3D pour l'analyse immersive nécessitent davantage d'efforts mentaux et physiques, il est possible de les mesurer à l'aide des réponses à la NASA-TLX brute.

H5 : *3DHelix* a la même facilité d'utilisation que *2DSpirals*, ce qui est mesuré par les réponses à l'échelle d'utilisabilité du système. Les interfaces de RV et leurs interactions étaient suffisamment simples et intuitives pour avoir une utilisabilité au moins aussi élevée que celle de la 2D.

H6 : *3DHelix* induit plus de flux que *2DSpirals*, ce qui est mesuré par les réponses à la Short Flow Scale 2. En effet, le prototype 3D utilise une technique de visualisation et des interactions qui permettent de se concentrer uniquement sur la tâche à accomplir, ce qui peut faciliter l'atteinte d'un état de flux.

4.4.4 Résultats

Résultats quantitatifs

Tous les participants ont détecté la rupture de périodicité sans alerte dans les deux conditions. Nous avons vérifié la normalité de la distribution des temps de réponse avec un test de Shapiro-Wilk (p = 3.7e-06) afin d'utiliser un test de Student pour comparer les temps de réponse moyens des deux conditions. Ce test de Student rejette l'hypothèse de différence de moyenne entre les spirales et les Cybercopters (p = 0,155).

Modality\Task	FN	A1	A2	A3	A4	A5
Cybercopter	100%	100%	80 %	75 %	55 %	100%
Spirals visualization	100%	96%	85%	70 %	50 %	100%

Table 4.2 – Taux de succès pour la classification des alertes

Les taux de réussite étaient presque similaires pour chaque alerte entre les deux conditions (tableau 4.2). La première alerte était la plus facile à détecter, car la rupture de périodicité est très claire, tous les participants l'ont trouvée avec les Cybercopters et un seul l'a manquée avec les spirales. La deuxième alerte suit de près la première et a été manquée par un plus grand nombre de personnes que la première (taux de réussite de 80

% pour les Cybercopters et de 85 % pour les spirales). Entre les deux premières alertes et la troisième, il y a un changement de périodicité dans les données, la périodicité du motif passe de 1920 s à 4400 s pour les deux capteurs, pour DPIT301, le motif peut également être trouvé avec une période de 2200 s comme les "deux pics" peuvent être interprétés comme "un court puis un long" (voir figure figure 4.10 c et c') du participant. Le taux de réussite de la troisième alerte était inférieur à celui des deux précédentes, avec 75 % pour les Cybercopters et 70 % pour les spirales. La quatrième alerte, qui est un faux positif, a eu le pire taux de réussite avec 55 % pour les Cybercopters et 50 % pour les spirales. Bien que les participants aient parfois été incapables de trouver une rupture des schémas périodiques qu'ils avaient trouvés après la troisième alerte, ils ont estimé qu'il était "plus sûr" de déclencher une alarme en cas de doute. Enfin, dans les deux conditions, tous les participants ont été en mesure de caractériser correctement la dernière alerte, ce qui correspond à une augmentation de DPIT301.

Questions	Cybercopter	Spirals
Q1. Determiner le moment auquel la menace est apparue	19/20	19/19
Q2. Déterminer la période d'un capteur avec une période d'un seul pic	19/19	20/20
Q3. Déterminer la période d'un capteur avec une période avec plusieurs pics.	7/9	6/8

TABLE 4.3 – Taux de réussite des réponses sur les caractéristiques des alertes. Pour la troisième question, uniquement DPIT301 (*set1*) présente une période avec plusieurs pics.

Parmi les participants qui ont détecté la première alerte, un seul n'a pas trouvé l'heure de la menace avec les Cybercopters, alors que tous ont pu le faire avec la visualisation des spirales. Dans les deux cas, ils ont tous donné la bonne réponse pour la périodicité du motif avant qu'il ne se brise. Pour la troisième question, les participants ont parfois trouvé la période du "double pic" de DPIT301 (4400 s) et parfois la période du "motif alternatif" (2200 s). Lorsqu'ils ont trouvé la première, nous leur avons demandé s'ils pouvaient identifier une période plus courte. Dans ce cas, ils ont tous trouvé la période de 2200 s. En utilisant les Cybercopters, 7/9 participants ont trouvé la bonne période, tandis qu'en utilisant les spirales, 6/8 l'ont trouvée (tableau 4.3).

La durée moyenne de l'expérience a été de 1 heure par participant, y compris l'accueil et les questionnaires. Le temps mesuré pour chaque ensemble représente le temps écoulé entre le début de l'expérience et la réponse à la dernière alerte. Nous avons utilisé le test

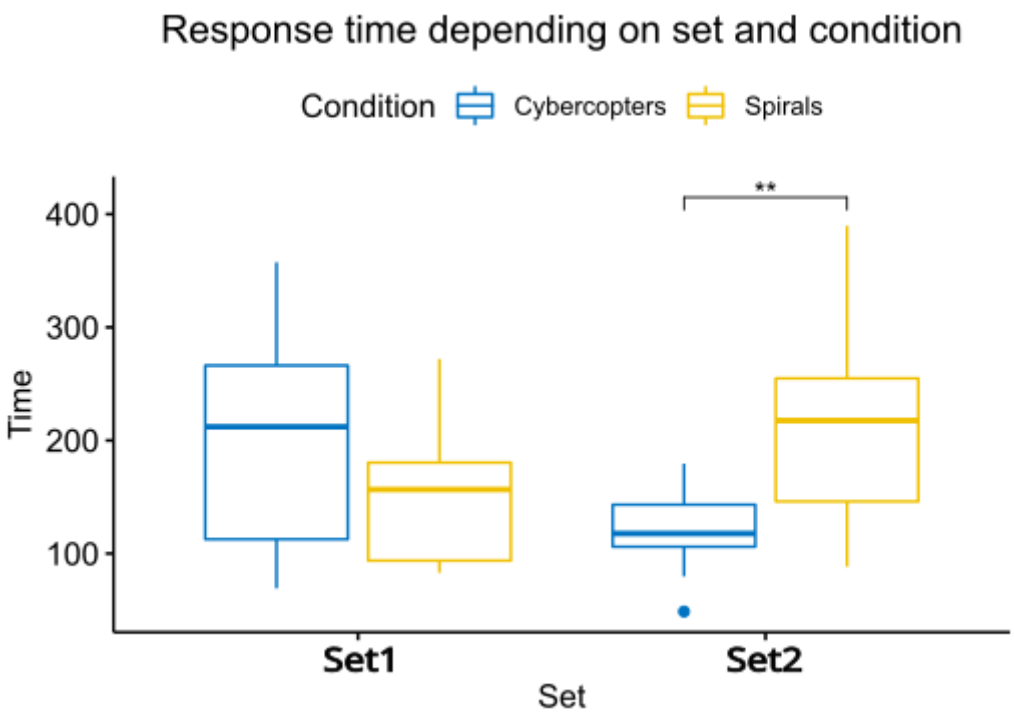


FIGURE 4.11 – Temps de réponse en fonction du *set* et de l'interface. Le temps de réponse à l'ensemble2 est significativement différent entre les deux interfaces ($p < 0.01$).

t corrigé avec la méthode de Holm-Bonferroni (figure 4.11) pour déterminer la différence statistique entre les conditions. Les Cybercopters sont significativement plus rapides pour *set2* (pvalue < 0.01) mais il n'y a pas de différence significative entre les Cybercopters et les spirales pour *set1* (pvalue $= 0.28$). Pour *set1*, le temps de réalisation moyen avec les Cybercopters est de 199 s (std = 99), avec les spirales, il est de 162 s (std = 61). Pour *set2*, le temps de réalisation moyen avec les Cybercopters est de 134 s (std = 58), avec les spirales, il est de 217 s (std = 90).

Pour mieux comprendre les différences de temps entre les deux ensembles, nous avons étudié le temps mesuré pour chaque alerte. Il représente le temps de réponse à l'alerte moins le temps de réponse à l'alerte précédente. Ainsi, les différences potentielles entre les temps de réponse des premières alertes ne biaisent pas les résultats des suivantes. Nous avons ensuite construit un modèle linéaire mixte avec le temps comme variable fixe, et les variables explicatives étaient : les conditions, les ensembles et les interactions entre les deux. Les conditions, ainsi que les interactions entre les conditions et les ensembles, sont significatives (pvalue = 0,009). Pour déterminer quelles paires d'alertes sont significativement différentes, nous avons utilisé le test t par paires corrigé par la méthode

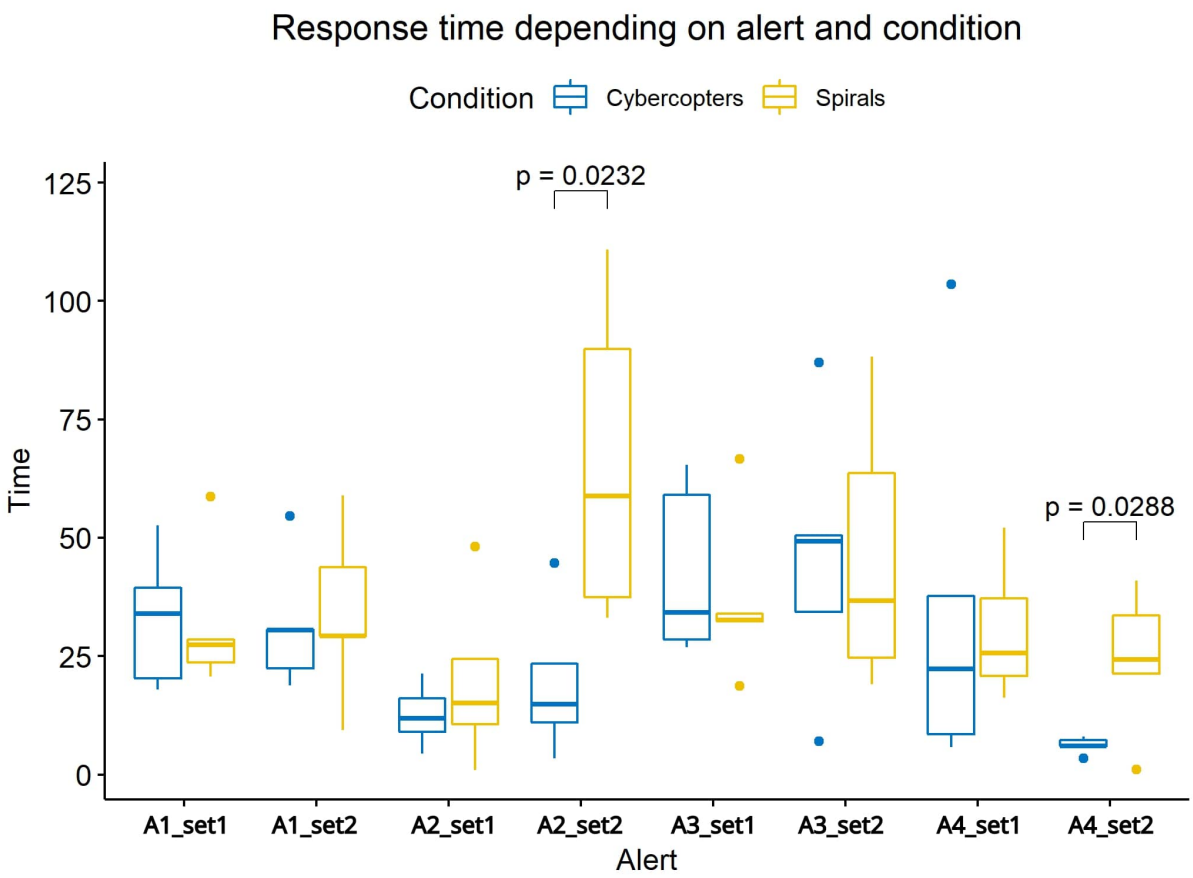


FIGURE 4.12 – Temps de réponse à chaque alerte en fonction du *set* et de l'interface. Les temps de réponse aux alertes 2 et 4 de *set 2* sont significativement différents entre les deux interfaces

de Holm-Bonferroni (figure 4.12). Les Cybercopters sont plus rapides pour les alertes 2 (valeur p = 0,02) et 4 (valeur p = 0,03) du *set2*. Dans tous les autres cas, il n'y a pas de différence de temps de réponse en fonction des conditions et des ensembles. Avec les Cybercopters, le temps de réponse moyen pour l'alerte 2 est de 19 s (std = 16) et de 6 s (std = 2) pour l'alerte 4. Avec les spirales, le temps de réponse moyen est de 66 s (std = 34) pour l'alerte 2 et de 24 s pour l'alerte 4 (std = 15).

Résultats quantitatifs

Questionnaires

[140] suggèrent qu'il est valable d'évaluer la différence de signification entre les réponses de chaque question d'un questionnaire à échelle de Likert : ils proposent d'utiliser une

ANOVA multiple à deux voies pour déterminer la signification de la différence entre deux conditions, c'est ce que nous avons fait dans notre analyse. Pour toutes les questions, à l'exception de la troisième de l'échelle de flux court, les seules différences significatives observées concernent les Cybercopters et le prototype en spirale.

Le score SUS est de 74 (avec une variation de 14) pour le prototype 2D et de 77 (avec une variation de 15) pour le prototype 3D. Les deux prototypes ont un score supérieur à 71, ce qui est généralement considéré comme une bonne utilisabilité ([141]). Un test de Shapiro indique que les données ne suivent pas une distribution normale (p=0,06), nous avons donc utilisé un test de Wilcoxon pour déterminer que la différence de facilité d'utilisation entre les deux prototypes n'est pas significative (p=0,45).

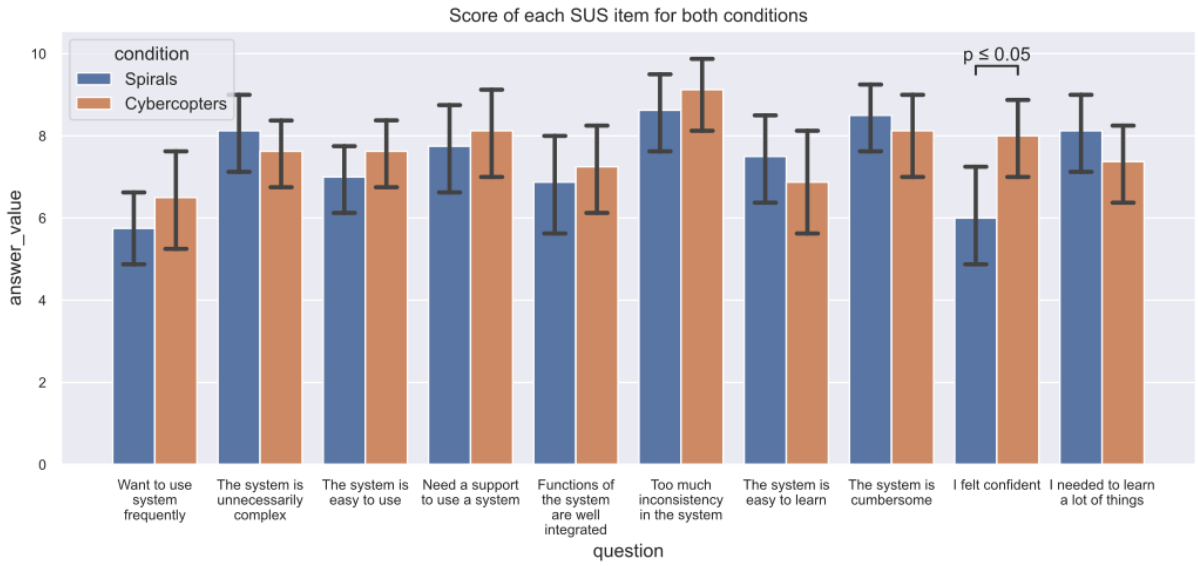


FIGURE 4.13 – Réponses détaillées aux SUS pour les deux interfaces. La seule différence significative entre les deux interfaces concerne la neuvième question, qui porte sur la confiance des participants.

La seule différence significative dans les items du SUS est pour le 9ème (pvalue = 0.012) : les participants se sentent plus confiants en 3D qu'en 2D (figure 4.13). (1.p=0.4, 2.p=0.2, 3.p=0.3, 4.p=0.4, 5.p=0.6, 6.p=0.4, 7.p=0.4, 8.p=0.6, 10.p=0.2) .

Les deux prototypes obtiennent un score de 27 au TLX de la NASA (avec une variation de 7 pour les deux), ce qui est considéré comme une charge de travail moyenne ([142]). La seule différence significative dans les items concerne l'exigence physique (p = .011) : les participants ont trouvé le Cybercopter plus exigeant physiquement que l'interface 2D (figure 4.14). (1.p=0.7, 3.p=0.5, 4.p=0.1, 5.p=0.5, 6.p=0.8)

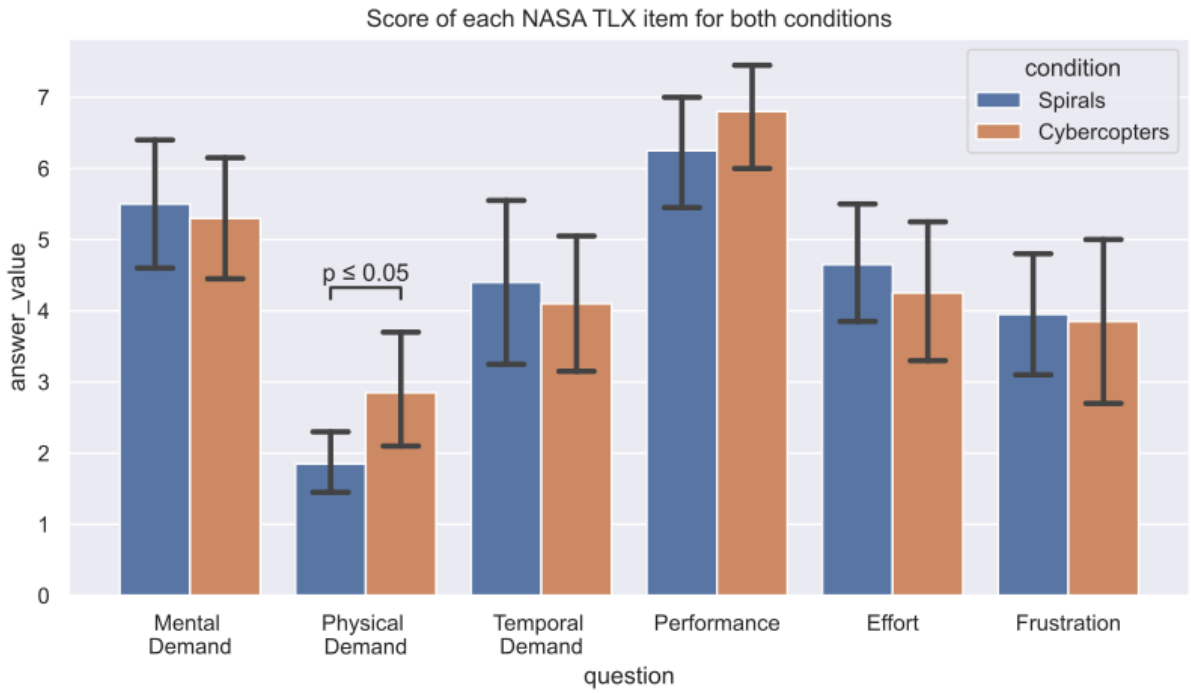


FIGURE 4.14 – Réponses détaillées au NASA-TLX pour les deux interfaces. La seule différence significative entre les deux interfaces concerne la deuxième question, qui porte sur les exigences physiques de la tâche.

Le flow moyen ressenti avec le prototype en spirale est de 3,42 (avec une erreur moyenne de 0,56) et le flux moyen ressenti avec le Cybercopter est de 3,74 (avec une erreur moyenne de 0,65). Comme le test de Shapiro indique que les données ne suivent pas une distribution normale (valeur p = 0,24), nous avons utilisé un test de Wilcoxon qui indique que la différence de flux induite chez les participants par les deux prototypes est significative (valeur p = 0,008). Les participants ont ressenti un état de compétence optimale légèrement plus élevé avec le prototype de réalité virtuelle. L'étude d'items spécifiques permet de mieux comprendre cette différence.

Pour les éléments du SHORT Flow State Scale (figure 4.15), les participants ont eu l'impression de mieux contrôler ce qu'ils faisaient avec le Cybercopter (valeur p = 0,025), ils ont trouvé que le temps passait différemment dans l'environnement virtuel (valeur p = 0,03), et ils ont trouvé l'expérience plus gratifiante avec l'interface immersive (valeur p <1e-05). Il y a eu une petite différence dans le sentiment de compétence des participants (pvalue = 0,07), ils ont eu l'impression d'être plus performants avec le Cybercopter. La question 3 de la Short Flow Scale est la seule qui dépende de l'interaction entre l'ensemble

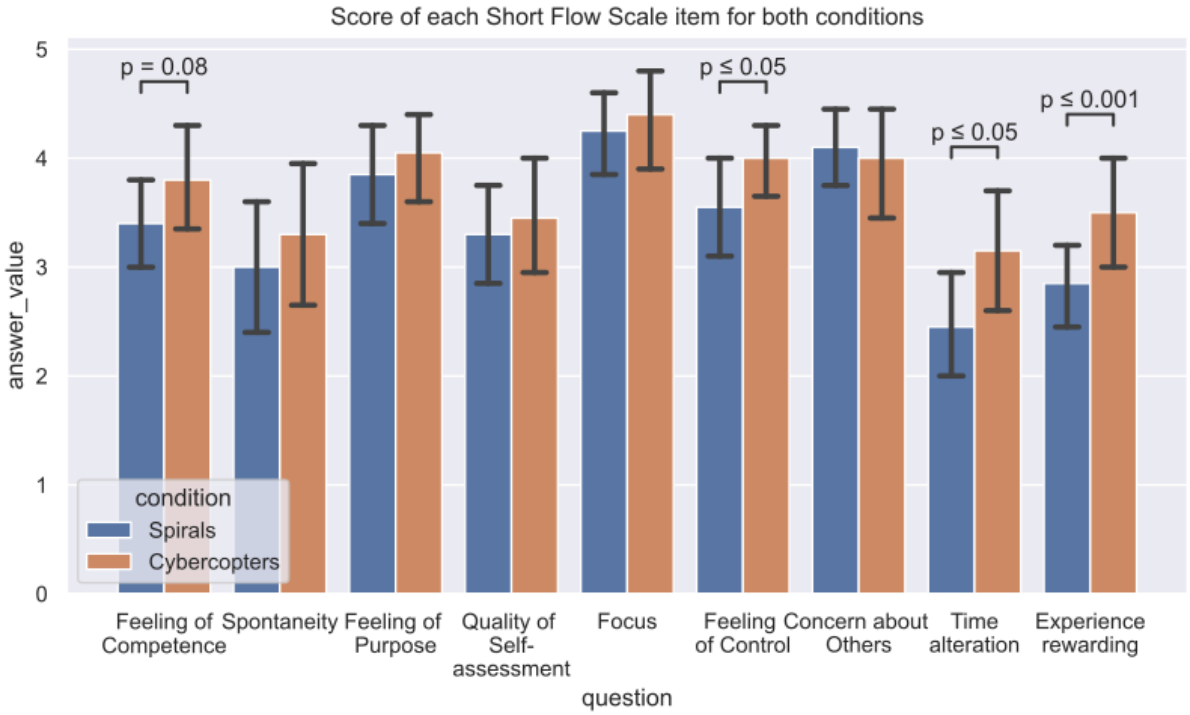


FIGURE 4.15 – Réponses détaillées au Short FLOW Scale pour les deux interfaces. Il existe des différences significatives pour les questions relatives au sentiment de compétence, aux contrôles, à la distorsion temporelle et au caractère gratifiant de l'expérience.

et la condition (valeur p = 0,002), nous avons donc procédé à une HSD de Tuckey pour savoir quelles interactions sont significatives. Lors de l'étude du *set2*, les participants se sont sentis plus spontanés avec le Cybercopter qu'avec les spirales.

Discussions libres et observations

Tous les participants ont apprécié l'utilisation de la solution immersive. Ils ont tous compris les deux métaphores après les tutoriels et tous les participants ont été capables d'utiliser le prototype immersif après le tutoriel de 10 minutes, même ceux qui n'avaient jamais utilisé la RV auparavant. Pour les deux conditions, les participants ont trouvé qu'il y avait un manque de points de référence, tels que les graduations le long de la ligne du temps. Huit participants n'ont pas utilisé le clic pour mettre en évidence les alertes et ont préféré utiliser les curseurs de temps pour détecter l'emplacement de l'alerte. Cinq participants ont préféré Cybercopter parce que les alertes sont à côté des données du capteur, deux parce que l'alignement est linéaire et non radial. Cinq participants ont trouvé que

set 2 était plus difficile que *set 1* (trois d'entre eux ont utilisé Cybercopter avec *set 2* tandis que deux d'entre eux ont utilisé des spirales), deux ont trouvé que *set 1* était plus difficile que *set 2* (ils ont tous deux utilisé des spirales avec *set 1*, ce qui pourrait indiquer que l'utilisation de spirales est plus difficile que l'utilisation de Cybercopters), d'autres n'ont pas trouvé de différence. Pour le prototype 3D, les participants ont particulièrement apprécié le joystick pour interagir avec la périodicité, car ils l'ont trouvé très naturel, même ceux qui n'avaient aucune expérience des manettes de jeu. Aucun participant n'a rencontré de problème d'occlusion ni n'en a signalé. En 2D, les participants ont affirmé qu'ils aimaient utiliser la molette, bien qu'ils ne l'aient jamais utilisée après le tutoriel.

Interviews des experts

Trois experts ont été interrogés, leur expérience allant de 3 à 15 ans. Tout ce que l'on peut dire de leur activité, c'est qu'ils ont tous travaillé dans des SOCs. Leurs avis divergent sur la réalité virtuelle, l'un d'entre eux utilisant quotidiennement son casque de réalité virtuelle pour gérer les différentes fenêtres de sa console d'outils dans un environnement 3D. Un autre est plus dubitatif, car la réalité virtuelle a le défaut de couper l'utilisateur du monde réel et donc de ses collègues, même s'il est très intéressé par les capacités de visualisation et d'organisation offertes par les environnements 3D. Il préférerait des représentations 3D en réalité augmentée, qui garderaient le lien avec le monde réel. Tous sont intéressés par les nouvelles interactions offertes par les casques de réalité virtuelle. Plus particulièrement, pour les Cybercopters, ils ont apprécié les possibilités de corrélations de différentes données avec les hélices. Ils auraient apprécié davantage d'interactions et d'options de filtrage avec les hélices et les spirales, par exemple la possibilité de réduire de moitié ou de doubler la période à l'aide d'un bouton ou d'aligner manuellement certains points spécifiques pour voir si un motif apparaît. Ils ont estimé que l'essaim d'hélices était intéressant, car il leur permettait de comparer plusieurs capteurs les uns par rapport aux autres. Ils ont également utilisé l'espace disponible pour trier les capteurs par groupe de similitude, en choisissant par exemple de sélectionner les capteurs qui affichent les mêmes signaux qu'un capteur suspect. Cependant, ils aimeraient utiliser l'espace virtuel pour transmettre davantage d'informations. Par exemple, l'espace 3D pourrait être utilisé pour représenter le plan de la station d'épuration ou d'autres graphiques (2D ou 3D) pour transmettre le contexte, comme le trafic internet pour détecter le chemin emprunté par le cyberattaquant pour perturber la station. Ils ont trouvé les interactions intuitives, mais l'un d'entre eux a suggéré d'utiliser une manette de jeu pour contrôler la représentation,

étant donné qu'ils étaient plus familiers avec ce type d'interacteurs. Le principal inconvénient relevé par les experts est que la réalité virtuelle les coupe de leurs outils habituels. Ils estiment qu'ils ont besoin d'un autre moyen d'accéder à leurs outils, tels qu'un collecteur de données, un IDS ou même des outils de communication classiques.

4.4.5 Interprétation des résultats

Les pourcentages de réponses aux alertes et de faux négatifs valident les exigences 2, 3 et 4 pour les deux systèmes de visualisation, Cybercopters et spirale, et valident H1. En outre, il n'y a pas de différence significative entre les pourcentages de réussite des deux systèmes de visualisation. Par ailleurs, les participants sont capables de détecter la première occurrence d'une menace, d'en donner la date ainsi que la période associée (simple ou double) pour les deux prototypes, ce qui valide l'exigence 1. Enfin, sur les 24 participants, 20 ont répondu avoir très peu d'expérience en cybersécurité ($<3/5$), ce qui valide l'exigence 5. Les deux prototypes sont donc efficaces pour répondre au problème posé, ce qui rejoint les résultats précédents de Lohfink et confirme l'intérêt de la visualisation 3D avec les Cybercopters. En ce qui concerne la mesure du temps de réponse, il n'y a pas de différence significative entre les deux prototypes pour la réponse faussement négative, ce qui valide H2 selon laquelle la représentation hélicoïdale est aussi efficace que les spirales pour détecter les ruptures de motifs périodiques pour cette expérience. Bien que nous n'ayons pas démontré de différence d'efficacité dans cette expérience, il est peut-être possible de le faire dans d'autres conditions. Ainsi, nous ne pouvons pas conclure que les spirales et les hélices sont similaires pour la détection de signaux périodique.

En ce qui concerne H3, le temps de réponse dépend non seulement de l'alerte traitée, mais aussi des capteurs mis en place, ce à quoi nous ne nous attendions pas. En effet, MV301 (*set2*) étant un actionneur du processus contrôlant DPIT301 (*set1*) leurs comportements semblaient suffisamment similaires pour être comparés. Il s'avère que ce n'est pas le cas, les Cybercopters aident à répondre à *set2* plus rapidement que les spirales, plus précisément, les Cybercopters aident à répondre aux alertes 2 et 4 significativement plus rapidement que la spirale. Par conséquent, nous rejetons l'hypothèse H3, car les Cybercopters ne sont pas plus efficaces que le prototype 2D pour répondre à toutes les alertes. Il est toutefois intéressant d'examiner les alertes pour lesquelles les Cybercopters permettent de répondre plus rapidement que les spirales. Dans ces cas, il est plus compliqué de distinguer une alerte d'une autre. En effet, les motifs de MV301 sont plus courts que ceux de DPIT301 et les temps d'alerte sont plus longs, peut-être qu'une combinaison des

deux rend la solution 2D moins efficace que la solution 3D parce que la représentation de l'alerte ne se superpose pas aux données. Ceci est cohérent avec le comportement des utilisateurs qui ont des difficultés à distinguer les différentes alertes lorsqu'il y a plusieurs alertes affichées sur la même spirale. En effet, les alertes 2 et 4 surviennent toutes deux peu de temps après l'alerte précédente (figure 4.10), ce qui a obligé les utilisateurs à jouer avec la taille de la fenêtre temporelle pour n'afficher qu'une seule alerte (il semble que la mise en évidence n'ait pas beaucoup aidé à séparer les alertes). Mais l'affichage d'une seule alerte masque son contexte, ce qui rend sa catégorisation plus difficile. Alors qu'avec les représentations hélicoïdales où le signal d'alerte est à côté des données, les participants n'ont aucun problème à distinguer une alerte d'une autre parce qu'elles sont à des hauteurs différentes. Ceci est confirmé par la différence de spontanéité entre les deux conditions sur *set2* rapportée dans le questionnaire de flux. La séparation entre les représentations des alertes et les données des capteurs facilite l'examen des données des capteurs tout en permettant de savoir où se produit l'anomalie. Elle confirme l'intérêt de la séparation entre les représentations tout en utilisant l'espace pour les corréler.

Les autres réponses aux questionnaires dépendaient uniquement de la condition (hélices ou spirales). Le score SUS des deux solutions confirme l'utilisabilité des deux prototypes, ce qui est cohérent avec les résultats de Lohfink et al. qui ont également trouvé que le prototype 2D était utilisable à l'aide du questionnaire ISO 9241/10 et soutient l'utilité des Cybercopters. Comme les deux interfaces ont une facilité d'utilisation comparable, nous validons H5.

Nous n'avons pas réussi à mettre en évidence une différence significative de charge de travail induite par les Cybercopters et les spirales, ce qui rejette H4. Cependant, les participants ont eu l'impression que les Cybercopters nécessitaient plus d'efforts physiques que l'interface 2D, ce qui est cohérent avec des travaux antérieurs [135]. Les participants se sont sentis significativement plus confiants (item 8 du SUS) et en contrôle (item 8 du Flow) avec les Cybercopters qu'avec la solution 2D. De plus, la représentation immersive donne plus de fluidité aux participants, ce qui valide H6 : elle donne un sentiment de compétence plus fort et le temps semble passer plus vite dans l'environnement 3D, peut-être en raison des interactions plus naturelles ou d'une meilleure immersion dans les données. La vitesse à laquelle les participants ont adopté les joysticks leur a permis d'interagir de manière plus fluide, ce qui leur a permis de rester concentrés sur les tâches à accomplir. Les participants ont trouvé Cybercopters plus gratifiant, probablement en raison de son aspect jeu vidéo, qui peut avoir un effet positif sur la motivation.

Enfin, comme l'a montré ([143]), les spirales et les hélices sont compatibles et peuvent être combinées pour obtenir le meilleur des deux mondes : la facilité d'utilisation d'une interface WIMP (Windows, Icons, Menus and Pointing device) et un vaste environnement pour corréler les données. Par exemple, on peut choisir d'utiliser une spirale pour consulter des données périodiques et passer à une hélice pour répondre à des alertes.

4.5 Implications pour la conception de systèmes d'Immersive Annalytics

Sur la base de notre expérience et de nos résultats, nous tirons des conclusions sur l'intérêt de l'Immersive Analytics par rapport aux visualisations 2D et plus spécifiquement dans le domaine de la cybersécurité :

- Si le but d'une représentation est de corréler plusieurs visualisations ensemble, alors même si une visualisation 3D n'est pas plus efficace qu'une visualisation 2D pour une tâche spécifique (comme la détection de signaux périodiques), l'ajout d'une dimension peut en valoir la peine. Dans un domaine où la corrélation entre les données est importante, comme la cybersécurité, cela peut être un atout précieux.
- Les concepteurs d'analyses immersives doivent faire attention à la physicalité de leurs interactions, car l'Immersive Analytics peut être plus fatigante même en position assise. Ce point peut également être considéré comme une opportunité d'aider les employés sédentaires à bouger.
- Les concepteurs d'outils d'analyse immersive devraient choisir avec soin les interacteurs et les interactions qu'ils proposent. Le fait que les participants choisissent d'interagir avec l'hélice uniquement à l'aide de joysticks plutôt qu'avec le curseur montre qu'un petit changement dans le choix de l'interacteur peut avoir un impact considérable. Des interacteurs spécifiques pourraient être développés pour chaque tâche de visualisation.
- Une meilleure visualisation et une meilleure interaction facilitent les tâches pour les participants, ce qui induit une plus grande fluidité chez les utilisateurs d'Immersive Analytics. Dans un domaine comme la cybersécurité, où les journées peuvent être longues et répétitives, cela pourrait permettre aux opérateurs d'être plus attentifs pendant la journée. Cependant, l'effet de distorsion temporelle pourrait être problématique dans un environnement sensible au temps où les décisions doivent être prises rapidement.

— Enfin, le plus grand inconvénient constaté par les experts est que l'Immersive Analytics les coupe de leur outil habituel. Pour remédier à ce problème, [144] a utilisé un casque de réalité augmentée afin que les utilisateurs puissent utiliser des visualisations en 3D tout en interagissant avec l'écran, le clavier et la souris de l'ordinateur. Une autre option pour permettre aux opérateurs de cybersécurité d'accéder à leurs outils tout en utilisant l'analyse immersive consiste à intégrer des fenêtres 2D dans un environnement 3D ([114]). Les deux solutions ont leurs avantages : la réalité augmentée permet aux opérateurs d'interagir avec leur environnement, y compris avec leurs collègues. Cependant, la réalité virtuelle permet de créer un nouvel espace de travail indépendant des conditions réelles, ce qui permet d'arranger les visualisations dans l'espace à volonté. Dans les deux cas, il pourrait être intéressant d'associer les outils classiques et l'analyse immersive pour soutenir un flux de travail complet. Nous explorons cette option dans le chapitre suivant.

4.6 Conclusion

Dans notre recherche, nous avons introduit une nouvelle approche utilisant des visualisations 3D pour classer les alertes dans des données périodiques en représentant divers attributs sur des visualisations séparées et en les corrélant le long d'un axe commun. Cette stratégie atténue les interférences entre les variations d'attributs et permet d'utiliser des visualisations sur mesure pour chaque attribut en fonction de ses caractéristiques inhérentes. Nous avons appliqué cette technique à la cybersécurité en créant une visualisation immersive en 3D utilisant des représentations hélicoïdales. Ce système discerne efficacement les motifs périodiques et les anomalies dans les données temporelles tout en s'adaptant à divers types d'informations. Bien que nous ayons corrélé deux attributs dans notre étude, il est possible d'inclure d'autres représentations autour des hélices. Dans le cadre d'une étude utilisateurs impliquant des experts en cybersécurité et des profanes, nous avons comparé notre prototype à une visualisation de l'état de l'art utilisant des représentations en spirales. Les participants ont évalué les alertes déclenchées par un algorithme de reconnaissance de motifs utilisant les données d'une usine de traitement d'eau subissant des cyberattaques. Nos résultats n'indiquent pas de différence significative entre les représentations hélicoïdales et les spirales, sauf pour certains scénarios où les hélices réduisent les temps de réponse des participants. En outre, les utilisateurs expriment une plus grande confiance et un plus grand engagement dans un état de flow avec la solution

immersive, malgré ses exigences physiques plus importantes. Malgré l'intérêt manifesté par les experts interrogés pour l'utilisation de l'analyse immersive, ils ont exprimé des réserves quant à l'impossibilité d'utiliser leurs outils de travail habituels lors de l'utilisation de notre prototype.

Résumé

Suite à note état de l'art sur l'usage de l'Immersive Analytics pour la cybersécurité, nous avons opté pour l'utilisation de représentations hélicoidales dans le but de visualiser des données périodiques et de les interconnecter avec d'autres représentations au sein d'un environnement 3D. Cette conception a été concrétisée à travers la création de deux prototypes : l'un, de nature ego-centrée et non immersif, et l'autre, basé sur des small multiples et exo-centré, offrant une expérience immersive. Nous avons procédé à l'évaluation de ces deux prototypes au moyen d'expériences utilisateur portant sur des cas d'utilisation distincts. Néanmoins, certaines conclusions se sont dégagées de nos analyses. Les avantages inhérents aux représentations hélicoidales résident dans leur capacité à faciliter la corrélation de ces représentations avec d'autres données, parfois de manière plus efficace qu'une représentation bidimensionnelle. Il semble donc plus approprié, dans le contexte de la cybersécurité, d'utiliser des small multiples exo-centrés pour la comparaison entre différentes représentations. Par ailleurs, nos expériences utilisateur ont révélé une expérience plus immersive, mais au prix d'un effort cognitif accru dans le cas des visualisations immersives. Enfin, les retours d'experts indiquent une ouverture à l'utilisation de la réalité virtuelle pour la cybersécurité. Cependant, il est souligné que l'emploi de ces visualisations ne devrait pas les isoler de leurs outils de travail habituels.

Chapitre 5

Intégration des outils métiers dans l'Immersive Analytics

Figure 5.1 – Vue d'artiste d'un environnement immersif et collaboratif pour la cybersécurité.

Dans le chapitre 2, nous avons établi que, pour être acceptées au sein de la sphère cybersécurité, les visualisations de données devaient être en mesure de prendre en charge l'ensemble du flux de travail des opérateurs de cybersécurité. Cela implique la capacité d'afficher non seulement les visualisations, mais également les systèmes de détection d'in-

trusion (IDS), les consoles d'outils, ainsi que l'ensemble des outils disponibles sur un ordinateur pour un opérateur. En ce qui concerne les interactions, les abstractions jouent un rôle crucial, de même que les sélections simples et multiples. L'annotation des données et la prise d'actions de protection sur le réseau sont également des éléments essentiels pour les opérateurs. Comme exposé dans le chapitre 3, un des défis de l'Immersive Analytics consiste à s'intégrer dans le contexte réel tout en permettant des interactions abstraites avec les données de cybersécurité. Il est nécessaire de développer ce type d'interaction pour que les systèmes immersifs puissent s'intégrer pleinement dans l'environnement des Centres Opérationnels de Sécurité (SOCs). Ces constatations sont corroborées par nos échanges avec des experts, comme cela a été mentionné dans le chapitre 4. Bien que ces experts soient ouverts à l'utilisation de la Réalité Virtuelle, ils ont exprimé le besoin d'accéder à leurs outils au sein de l'environnement virtuel.

Dans ce chapitre, nous décrivons en détail comment nous avons étudié la mise en place d'un espace de travail en Réalité Virtuelle dédié à la cybersécurité (Figure 5.1), intégrant à la fois les outils traditionnels de cybersécurité via des fenêtres 2D interactives et intégrées dans l'environnement virtuel ainsi que des visualisations de données 2D et 3D des représentations (Tableau 5.1). Nous expliquons également comment nous adaptons le pipeline de visualisation à cette tâche, ainsi que les problématiques techniques que cette réalisation entraîne. En effet, bien que chaque brique technique existe et peut théoriquement être reliée aux autres, leur couplage n'a jamais été réalisé auparavant. De plus, l'utilisation de plusieurs modes d'interaction pour manipuler à la fois des outils en 2D et des représentations en 3D soulève des questions quant à la transition entre ces modes et la manière optimale de la réaliser.

Enfin, nous présentons deux prototypes. Le code source de ces prototypes est disponible pour les partenaires industriels de la chaire Cyber CNI. Le premier prototype a été présenté dans le cadre de l'ECW 2022 pour illustrer l'intégration des outils en 2D dans un environnement virtuel dédié à la cybersécurité. Le second prototype vise à étudier les changements de paradigmes en matière d'interaction avec des représentations en 3D. L'objectif à plus long terme est de créer un environnement immersif dans lequel les travaux de recherche de la chaire Cyber CNI peuvent être menés pour contribuer à l'avancement de l'Immersive Analytics en cybersécurité.

Besoins	Réalisations
Assurer le transfert des données à partir du début de pipeline de visualisation (sources cyber) jusqu'à la fin (frameworks de visualisations immersives)	Agrégation et filtrage des données avec le SIEM ELK, puis transfert par fichier CSV au framework d'Immersive Analytics IATK
Interagir et modifier les données efficacement à toutes les étapes du pipeline de visualisation	Opération sur les jeux de données et paramétrage des représentations : clavier/souris + fenêtre 2D avec udekstopduplication Interactions avec les représentations de données : interactions native à IATK et Unity
Passer d'un paradigme à l'autre	Prototype multi paradigme : clavier magic keyboard + souris, mains libres avec leapmotion, manettes de casque oculus quest

TABLE 5.1

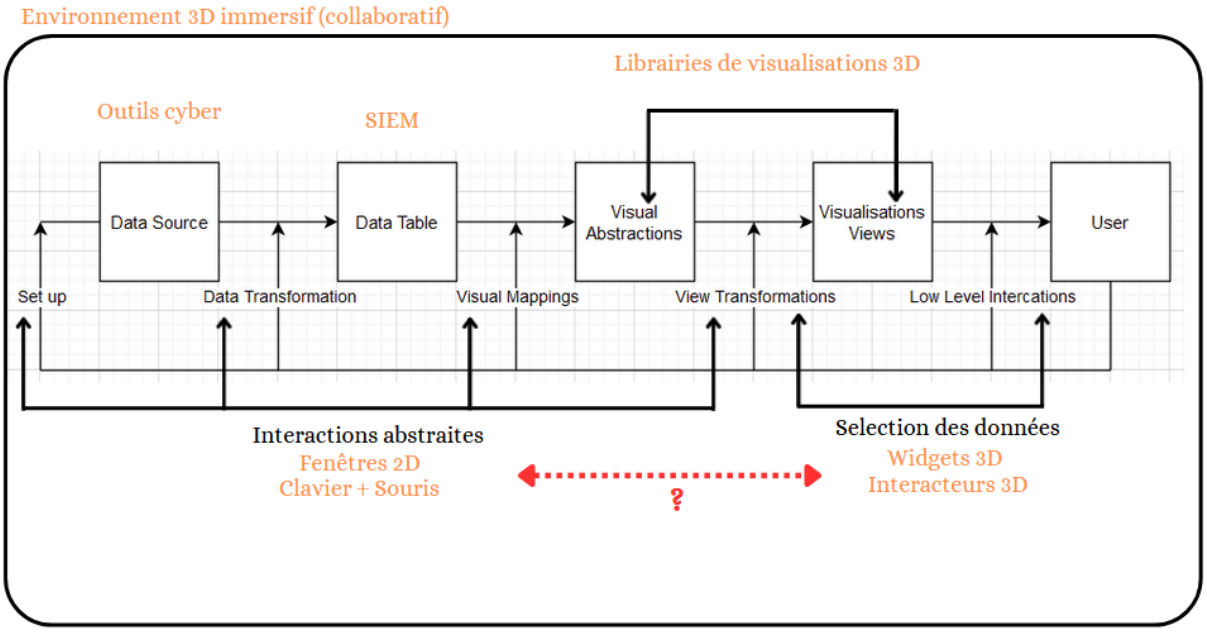


FIGURE 5.2 – Pipeline de la visualisation adapté pour l'Immersive Analytics

5.1 Concepts

Même si le fonctionnement global du pipeline de visualisation demeure inchangé dans l'environnement immersif, de nouvelles interactions sont nécessaires pour rendre les différentes visualisations utilisables. En ce qui concerne le flux de données (représenté par les boîtes en haut du diagramme Figure 5.2), les différentes composantes technologiques existantes permettent la création d'un pipeline de visualisation. En effet, le domaine de la cybersécurité est caractérisé par une multitude d'outils variés, allant du sniffeur de réseau à la gestion des pare-feux, ce qui engendre la gestion de nombreuses sources de données. Heureusement, cette problématique est bien connue en cybersécurité, et les Systèmes de Gestion de l'Information et des Événements de Sécurité (SIEM, pour Security Information and Event Management) permettent de convertir toutes ces sources en tables

de données exploitables. Il reste ensuite à trier ces tables en vue de leur affichage, et une fois de plus, les SIEM fournissent des outils à cet effet. Pour afficher ces données au moyen de bibliothèques de visualisation 3D, il suffit de leur fournir ces tables, avec des formats de fichiers couramment utilisés tels que le JSON ou le CSV qui peuvent parfaitement remplir cette fonction. D'un point de vue technique, la mise en place du flux de données semble théoriquement possible, étant donné que toutes les composantes nécessaires existent déjà. Toutefois, il convient de noter que la concrétisation d'un tel projet demeure jusqu'à présent inédite.

Une autre complexité réside dans les interactions pour l'utilisateur, qui doit être capable d'utiliser aussi bien les outils en 2D que ceux en 3D. En ce qui concerne les représentations 3D et les objets virtuels, il est envisageable d'utiliser des interactions immersives intégrées par le biais de bibliothèques existantes ou de concevoir nos propres solutions. En revanche, pour interagir avec les premiers outils 2D de la cybersécurité, il doit effectuer des actions abstraites, souvent en utilisant des lignes de commande, puis il doit passer à des interactions de filtrage et de sélection qui se dérouleront dans l'environnement en 3D. L'intégration des outils de cybersécurité classiques peut être réalisée en utilisant des fenêtres 2D intégrées à l'intérieur de l'environnement virtuel, ainsi que des interactions au clavier et à la souris. Il est possible de maintenir la position de l'écran physique dans le monde virtuel afin de ne pas désorienter l'utilisateur. La littérature propose certaines solutions pour interagir avec les écrans 2D dans les environnements virtuels, souvent basées sur des techniques de ray-casting ou des tablettes tactiles proches de l'utilisateur [114]. Bien qu'elles soient intuitives, elles restent imprécises et plus lentes que l'utilisation des clavier et souris réels [145]. Pour "apporter" le clavier et la souris dans le monde virtuel, Hoppe et al. [145] ont utilisé des capteurs afin de connaître leurs positions et les retranscrire dans le monde virtuel, ce qui permet d'interagir efficacement avec les fenêtres 2D. Cependant, le changement de paradigme entre interacteurs 2D et interacteurs 3D peut-être un problème pour les utilisateurs qui reste peu abordé. De plus, la littérature ne propose pas de recommandation sur les interacteurs 3D à utiliser conjointement au clavier/souris.

Ainsi, deux questions se posent : quelles composantes techniques doivent être employées pour connecter les outils de cybersécurité existants aux visualisations en réalité virtuelle ? Et comment permettre à l'utilisateur de passer d'un paradigme d'interaction à un autre sans perturber son travail ?

5.2 Immersive Cyber Workspace : choix techniques

La figure 5.3 illustre les différentes technologies envisagées pour implémenter le concept d'espace de travail immersif pour la cybersécurité.

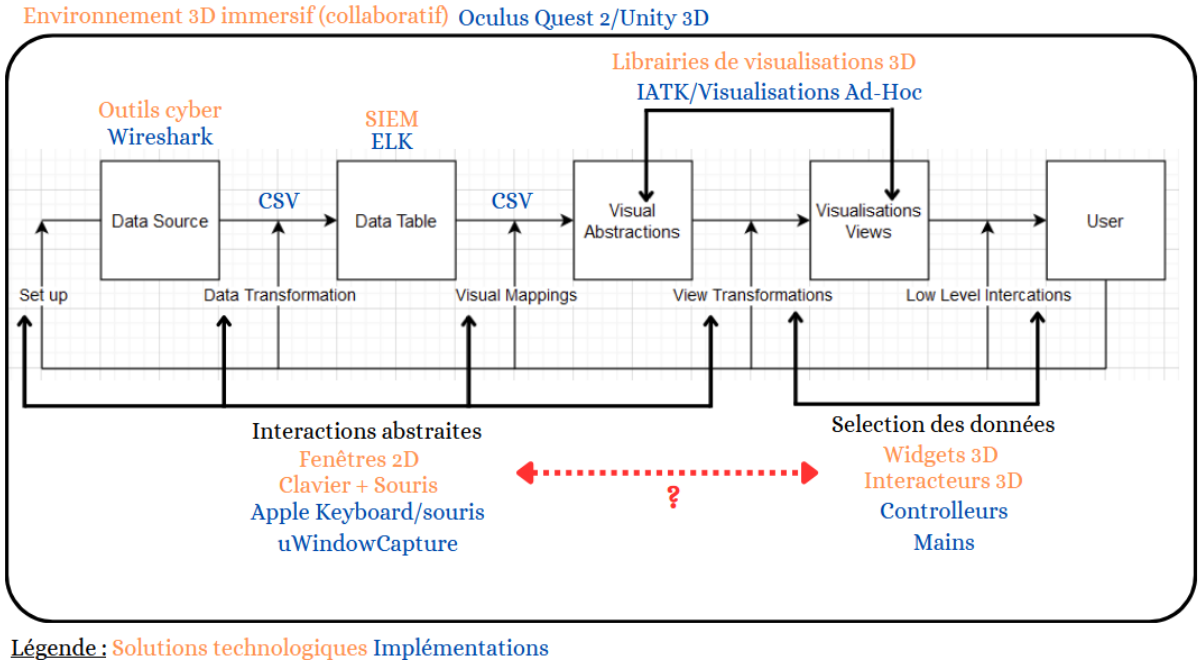


FIGURE 5.3 – Exemples des briques technologiques et logicielles pouvant servir à implémenter le pipeline de visualisation pour l'Immersive Analytics

En ce qui concerne la collecte et la gestion des données en cybersécurité, divers outils tels que Wireshark, Snort, ou les pare-feux sont envisageables. Dans le cadre d'expérimentations, il est également prévu d'utiliser des ensembles de données de la littérature, à l'instar du UNSW-NB15 [146]. Pour l'organisation et la collecte de ces données cybernétiques, la solution retenue est un SIEM, tel que la "stack" ELK, qui constitue un standard de l'industrie. En effet, grâce à Logstash, les données peuvent être structurées en tables manipulables, puis filtrées via Elasticsearch, et enfin visualisées grâce à Kibana, qui propose une interface graphique conviviale. Ainsi, nous prenons en charge les interactions abstraites, l'intégration, et le tri des données.

Concernant les solutions immersives, Unity est un moteur de jeu prenant en charge la 3D, notamment la réalité virtuelle, grâce à diverses bibliothèques qui gèrent les aspects techniques, y compris les interactions de base avec les objets 3D, comme la saisie d'objets ou l'activation de boutons. Cette flexibilité permet de connecter facilement différents types

de casques et d'interactions. Unity permet d'exécuter les applications de deux manières : directement depuis l'éditeur pour le développement, avec la possibilité de modifications en temps réel, ou sous forme d'un build, qui constitue la version définitive et compilée du jeu, spécifique à une plateforme donnée, prête à être distribuée et exécutée par les utilisateurs, offrant une efficacité et une optimisation supérieure par rapport au code d'édition. Il est à noter que les applications peuvent être déployées directement sur un casque autonome de RV si on ne souhaite pas utiliser un PC en phase d'exécution.

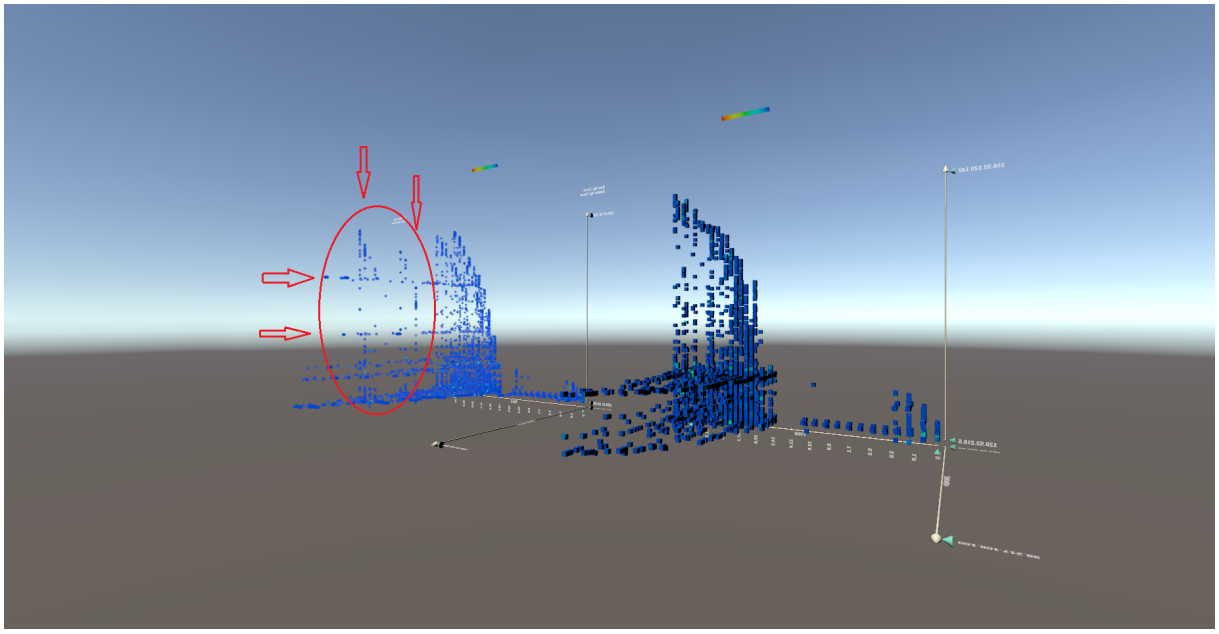


FIGURE 5.4 – Visualisation de deux périodes différentes du dataset UNSW-NB15 [146] avec IATK [113], à gauche sans attaque, à droite pendant une attaque. On peut remarquer que des ordinateurs jusque-là silencieux commencent à communiquer

Afin de réaliser automatiquement des visualisations à partir des données transférées par les outils de cybersécurité, nous utilisons la librairie IATK [147]. Des deux librairies compatibles avec Unity, IATK (Figure 5.4) est le seul capable d'afficher un grand nombre de données [148]. Il permet la génération automatique de visualisations complexes à partir de sources en CSV, tout en proposant une interface pour configurer les visualisations à l'aide d'une grammaire graphique depuis l'éditeur Unity (et donc utilisable uniquement dans la version en développement du prototype). IATK a également été adapté et utilisé dans RagRug [147], une solution de Situated Analytics, pour des capacités en temps réel et une intégration avec Node-RED, démontrant ainsi sa capacité à s'adapter à différents pipelines de visualisation. Nous avons utilisé IATK pour représenter certains jeux

de données cyber à l'aide de space-time cubes, qui permettent d'identifier l'apparition de patterns suspects (Figure 5.4). Par conséquent, notre cadre de visualisation immersive se concentre sur IATK en raison de ses capacités d'adaptation et de sa capacité à gérer de grandes quantités de données. De plus, pour étudier la transition d'un paradigme d'interactions en 2D vers un paradigme d'interactions immersives, il est nécessaire de réaliser des visualisations hautement configurables comportant des interactions spécifiques compatibles avec divers types d'interacteurs. En effet, bien que IATK soit efficace pour représenter automatiquement des données, elle présente des limitations en termes de souplesse, notamment en ce qui concerne les interactions et la modification des vues affichées en temps réel. Les bibliothèques de shaders et d'interactions offertes par Unity permettent de créer ce genre de système, comme nous l'avons fait dans le chapitre 4.

En ce qui concerne le choix du casque, nous optons pour l'Oculus Quest 2, car il permet non seulement de suivre le clavier dans l'environnement virtuel nativement, mais aussi d'afficher les mains de l'utilisateur lorsqu'il les place au-dessus du clavier pour une utilisation fluide et naturelle [1]. Enfin, pour intégrer les fenêtres 2D dans l'environnement virtuel, nous utilisons UWindowCapture d'Hecomi [2], car cette solution peut être intégrée à Unity et manipulée avec des plugins de Réalité Virtuelle. Ces fenêtres sont des répliques des fenêtres du bureau de l'utilisateur, ce qui lui permet d'interagir avec des techniques de raycasting ou avec la souris. Les plugins VR de Unity permettent d'interagir facilement avec les fenêtres 2D, permettant de reproduire les outils "macros" proposés par Hoppe et al. [114], qui permettent par exemple de zoomer et de dézoomer la fenêtre. UWindowCapture utilise les fonctions natives de Windows pour capturer en temps réel les fenêtres du bureau et les projeter sur des textures 2D dans l'environnement virtuel, chaque fenêtre pouvant être utilisée indépendamment.

5.3 Réalisations

Dans le cadre du défi "Centre de Conduite des Opérations Cyber Militaires 3.0" qui s'est tenu lors de l'ECW 2022 [3], nous avons développé un prototype mettant en œuvre certains des concepts présentés précédemment. L'objectif était de permettre à un commandement militaire cybernétique d'obtenir une vue d'ensemble complète du déroulement

1. https ://developer.oculus.com/documentation/unity/tk-ovr-tracked-keyboard-prefab/, 02/10/2023
2. https ://github.com/hecomi/uWindowCapture, 02/10/2023
3. https ://www.defense.gouv.fr/aid/actualites/gagnants-du-defi-centre-conduite-operations-cyber-militaires-30, 02/10/2023

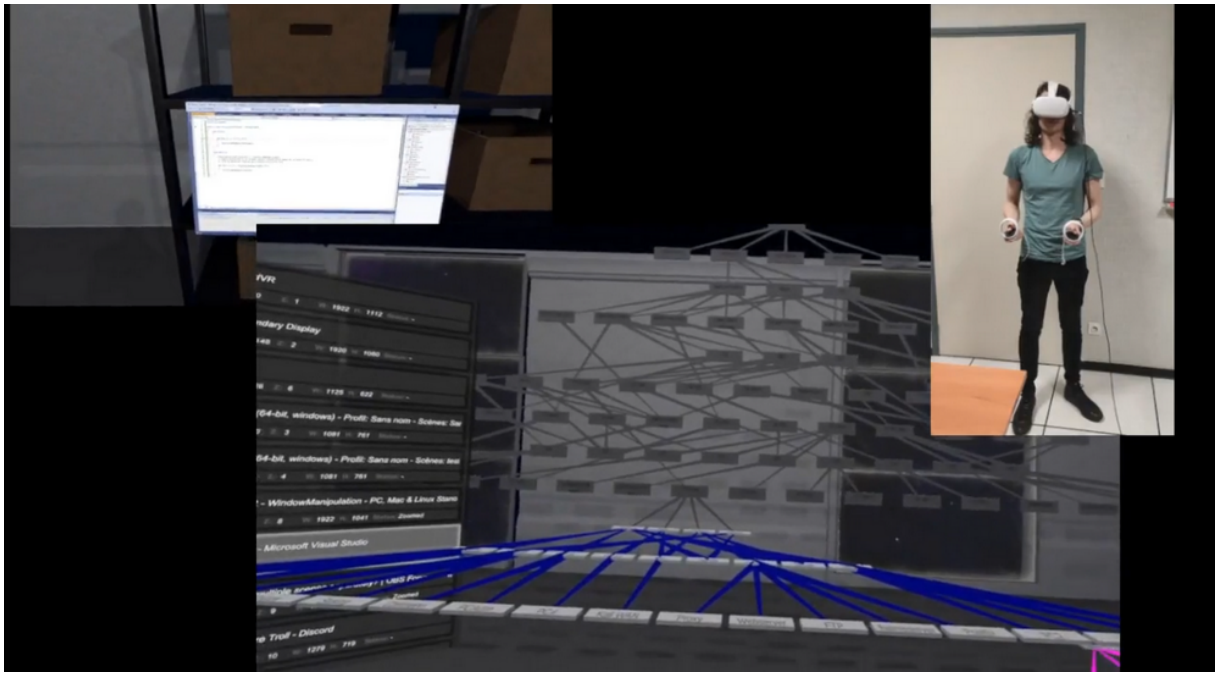


FIGURE 5.5 – Prototype présenté à l'ECW 2022. L'utilisateur (en haut à droite) peut accéder à une représentation sous forme de graphs de assets d'une mission (en bas) tout en ayant accès à ses outils habituels (en haut à gauche).

d'une opération. Pour ce faire, il était essentiel de disposer d'informations sur le déroulement de l'opération sur le plan opérationnel, ainsi que sur l'état des ressources à sa disposition pour accomplir les différentes tâches nécessaires à la réussite de l'opération. Nos partenaires industriels nous ont fourni un scénario fictif d'opération, similaire à ce que pourrait être un scénario réel, sous forme de graphe.

Nous avons choisi de représenter ce graphe (Figure 5.5) sous forme d'un graphe 2D vertical dans l'environnement virtuel, où chaque nœud représentait une tâche à accomplir pour mener à bien la mission. Lorsque l'utilisateur cliquait sur l'un des nœuds représentant une ressource, un diagramme horizontal se déployait, permettant d'accéder au réseau de cette ressource. En cas d'incident affectant une ressource (potentiellement dû à une attaque), le nœud correspondant changeait de couleur, et tous les nœuds dépendant de cette ressource changeaient également de couleur pour indiquer la compromission de la mission. De plus, les fenêtres 2D du bureau étaient accessibles dans l'environnement virtuel si l'opérateur avait besoin de ses outils habituels.

Malgré toutes les fonctionnalités que possédait ce prototype, qui combinait avec succès la visualisation 3D et les interfaces 2D dans l'environnement virtuel, il présentait des

limitations. Le scénario fictif ne permettait pas de relier les données de cybersécurité aux représentations graphiques comme souhaité. De plus, bien que le prototype fonctionnait correctement sur un ordinateur portable équipé d'une carte graphique, il était instable et souffrait de plantages inexpliqués. En outre, les fonctionnalités d'interaction pour le changement de visualisation graphique d'IATK n'étaient disponibles que dans le mode éditeur de Unity, c'est-à-dire sur l'écran du PC et pas dans l'environnement virtuel utilisé par l'utilisateur.

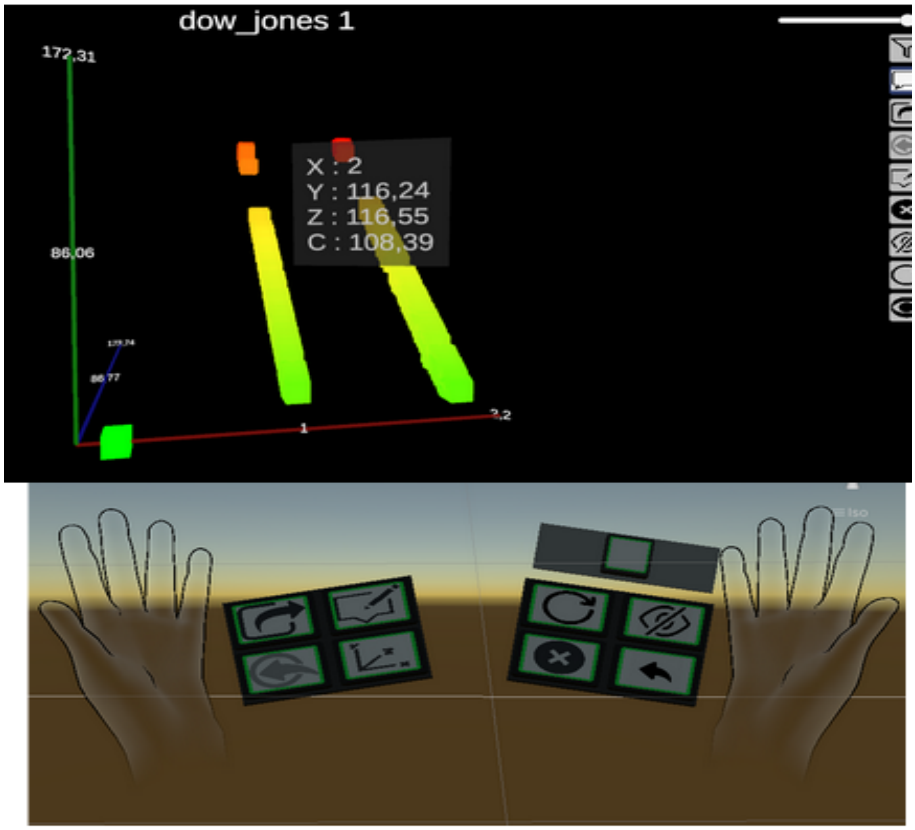


FIGURE 5.6 – Exemples d'interactions avec notre prototype. En haut, l'interface clavier/souris propose les boutons d'interactions à droite de l'écran. En bas, les mêmes interactions sont disponibles via des menus adjacents aux mains de l'utilisateur.

De plus, afin d'éprouver les divers paradigmes d'interactions accompagnés des représentations immersives et de faciliter la transition entre ces paradigmes, nous avons conçu un deuxième prototype. Celui-ci a la capacité de générer des visualisations permettant d'évaluer diverses interactions ainsi que le passage d'un paradigme d'interactions à un autre. Il permet l'affichage de nuages de points et de matrices cubiques à partir de données extraites d'un fichier CSV. Par ailleurs, il prend en charge les interactions au moyen du

clavier, de la souris, des mains libres grâce au Leap Motion [4], et des contrôleurs de casque de Réalité Virtuelle pour effectuer les interactions cruciales à la cybersécurité identifiées au chapitre 2, telles que le chargement et l'enregistrement de données, les annotations, les sélections, la manipulation des vues, les filtres, et la création de tooltips (Figure 5.6). Ainsi, nous disposons d'une capacité de visualisation pouvant s'associer simultanément à plusieurs interacteurs différents pour évaluer leurs performances. Ce prototype peut également être couplé à celui précédemment décrit afin de bénéficier des fenêtres 2D.

5.4 Conclusion

Le premier prototype élaboré pour l'ECW nécessite encore des améliorations en termes de stabilité, mais il démontre néanmoins les possibilités d'intégrations d'outils de cybersécurité dans l'environnement virtuel par le biais de fenêtres 2D. La mise en place d'une version construite (build) permet d'améliorer les performances, bien qu'elle ne résolve pas le problème des plantages. Cependant, il est important de noter que cette approche comporte une limitation, car elle supprime certaines options disponibles dans l'éditeur, notamment en ce qui concerne les interactions avec IATK et les autres objets 3D. En effet, de nombreux ajustements peuvent être facilement effectués dans l'interface 2D de l'éditeur.

Le deuxième prototype pourrait servir de base pour des expérimentations visant à comparer les changements de paradigmes lors de la transition entre les interacteurs au clavier. Les mesures envisagées incluraient le temps d'exécution, le flux de travail et l'utilisabilité. Nous soutenons l'idée que les interactions manuelles avec les représentations virtuelles favorisent une transition plus fluide vers les interactions au clavier, tout en perdant l'avantage des contrôleurs de pouvoir offrir davantage d'options d'interactions grâce à leurs boutons, notamment l'ouverture de menus et de widgets.

L'objectif réside dans la fusion de ces deux prototypes au sein d'un espace de travail expérimental unique. Cet espace bénéficierait à la fois de visualisations automatiquement générées pour fournir un contexte, et de visualisations personnalisées afin de tester de nouveaux concepts en matière de visualisation et d'interaction. Cette intégration permettrait d'établir un système permettant la visualisation de cas concrets de cyberattaques étudiés au sein de la Chaire Cyber CNI.

4. https ://www.ultraleap.com/leap-motion-controller-whats-included/, 02/10,2023

Résumé

Nous avons constaté que le pipeline de visualisation peut être adapté aux représentations immersives des données en cybersécurité, étant donné que toutes les composantes technologiques nécessaires sont disponibles. Cependant, l'utilisation de ce pipeline engendre des changements de paradigmes d'interactions inédits. Dans un premier temps, un prototype a été élaboré dans le but de rendre possible l'automatisation du pipeline de visualisation pour l'Immersive Analytics au sein de l'environnement virtuel. Ensuite, un autre prototype a été conçu afin d'étudier les changements de paradigmes d'interactions et leur impact sur le flux de travail d'un opérateur. Ces prototypes peuvent désormais être fusionnés pour créer une plate-forme de visualisation immersive pour la cybersécurité.

Conclusion et Perspectives

Conclusion

L'objectif de cette thèse consistait à examiner comment la Réalité Virtuelle, notamment l'Immersive Analytics, pouvait contribuer à la gestion des signaux faibles en cybersécurité, en se concentrant particulièrement sur les Centres Opérationnels de Sécurité (SOC) chargés de la défense en temps réel des infrastructures cyber. À travers l'étude du fonctionnement des SOCs effectué au chapitre 1, nous montrons qu'il est possible d'appliquer à la cybersécurité la notion de signaux faibles issue des études sur les comportements criminels, où la corrélation de plusieurs indicateurs suspects constitue souvent le meilleur indice d'un problème émergent. L'un de ces indicateurs spécifiques à la cybersécurité est la déviation du comportement du système surveillé par rapport au comportement nominal. Afin d'aider à la détection de ces indicateurs dangereux, des visualisations sont développées par la communauté scientifique, mais restent peu adoptées au sein des SOCs.

Dans le chapitre 2, nous utilisons la typologie de Munzner [3] qui permet de décrire les visualisations de la littérature en fonction du type de données visualisées, du type de tâches effectuées par l'opérateur, et des techniques de visualisations et d'interactions mises en place, nous a permis d'identifier les raisons de ce manque d'adoption. En effet, le processus opérationnel des opérateurs en cybersécurité est conforme aux principes du traitement des données massives de Van Ham et Perer [44] (Rechercher, afficher le contexte, développer à la demande). Ces opérations s'appuient principalement sur de simples visualisations courantes, intégrées dans des systèmes multivues qui requièrent une interaction substantielle. Les visualisations utilisées ont l'avantage d'être connues des opérateurs cybers, mais ont plusieurs défauts. En effet, nous avons constaté que les visualisations 2D éprouvent des difficultés à représenter les comportements temporels et à corréler efficacement les données, deux aspects cruciaux pour la détection des signaux faibles en cybersécurité. Par ailleurs, l'utilisation de l'espace de visualisation pour l'organisation des données et la collaboration ont été identifiées comme des pistes intéressantes.

Dans le chapitre 3, nous montrons que l'Immersive Analytics peut dépasser certaines des limitations des visualisations 2D, notamment en permettant d'accentuer les motifs

périodiques au sein des données, qu'elles soient continues ou liées aux réseaux, tout en facilitant la corrélation entre différentes visualisations dans l'espace 3D. Nous avons également constaté que le concept d'espace de visualisation pour organiser les données pouvait être amélioré en utilisant un environnement 3D immersif. En revanche, nous avons noté que ce genre d'environnement ne propose pas d'intégration efficace des outils 2D classiques comme ceux utilisés en cybersécurité.

Le concept de visualisation hélicoïdale que nous avons ensuite proposé dans le chapitre 4, le Cybercopter, permet de mettre en valeurs les éléments périodiques tout en les corrélant avec d'autres visualisations. Les résultats de nos études utilisateurs, menées avec des utilisateurs novices et experts, ont montré que les visualisations hélicoïdales sont plus efficaces que des visualisations 2D de l'état de l'art dans certains cas où les données sont complexes. De plus, les utilisateurs ont un usage du prototype immersif plus fluide, au prix d'efforts physiques plus élevés. Ce prototype nous a aussi permis de rencontrer des experts en cybersécurité pour avoir des retours non seulement sur le prototype, mais plus généralement sur la Réalité Virtuelle pour la cybersécurité.

Leurs retours ont mis en évidence la nécessité pour les opérateurs d'accéder à leurs outils cyber et au contexte des visualisations dans l'environnement virtuel, soulignant ainsi l'importance de créer un espace immersif intégrant à la fois des outils 2D et des visualisations 3D pour la cybersécurité. Dans le chapitre 5, nous décrivons les prototypes fonctionnels, mais limités techniquement, que nous avons réalisés pour explorer les interactions dans un tel environnement, qui nécessite des changements de paradigmes d'interactions entre 2D et 3D jusqu'alors inexplorés.

Perspectives

Les perspectives à ces travaux de thèse concernent la création de bonnes pratiques pour l'usage de visualisations immersives, afin de rendre l'Immersive Analytics utilisable en relevant le défi technique, et en mettant en œuvre les bonnes pratiques. D'un point de vue cybersécurité, elles cherchent à améliorer l'efficacité de la gestion des signaux faibles et, de manière plus générale, à optimiser le workflow des opérateurs grâce à l'utilisation de visualisations immersives.

Tout d'abord, nous envisagerons une exploration approfondie du concept de Cybercopter, en mettant l'accent tant sur les interactions avec la période temporelle des représentations que sur les outils permettant un filtrage intuitif des données. Il serait également

intéressant de comprendre de manière plus précise quelles caractéristiques des données confèrent à la représentation hélicoïdale une efficacité accrue : cela pourrait résulter du type de motif, de la taille et de l'emplacement des alertes, ou peut-être d'une combinaison de ces facteurs. Ainsi, il est nécessaire d'expérimenter davantage afin de répondre à ces questions et de contribuer à l'élaboration d'un espace de conception dédié aux visualisations hélicoïdales.

Créer un environnement de travail immersif dédié à la cybersécurité représente un défi technique qui nécessite l'intégration de divers composants logiciels. Cependant, cette entreprise est tout à fait réalisable et, à terme, pourrait servir de plateforme collaborative. Cet environnement ouvre des perspectives novatrices pour simplifier le traitement des signaux faibles par les utilisateurs, ce qui pourrait accroître la réceptivité des experts en cybersécurité envers l'Immersive Analytics, actuellement confrontés à la nécessité d'envisager son utilité sans disposer d'un prototype concret pour les guider. De plus, il permettrait la réalisation d'expériences cyber plus réalistes, en offrant des scénarios plus complets et authentiques, ce qui permettrait de tester de nouvelles représentations ad hoc de façon plus réaliste. Par ailleurs, il faciliterait l'étude de l'intérêt de l'organisation spatiale pour la création de sens tout en offrant la possibilité d'évaluer différents paradigmes d'interactions pour combler le fossé entre l'Immersive Analytics et les interfaces WIMP (Windows, Icons, Menus and Pointing device). À travers ces évaluations, il serait possible de proposer des recommandations en vue d'améliorer l'utilisabilité et l'efficacité des systèmes d'Immersive Analytics.

Chapitre 6

Contributions

6.1 Publications Scientifiques

6.1.1 Workshop

N. Delcombel, A. Kabil, T. Duval et al., « CyberCopter : a 3D helical visualisation for periodic signals of cyber attacks », in *VR4SEC*, 2021, p. 1-5

6.1.2 Journal

N. Delcombel, T. Duval et M. O. Pahl, « Cybercopters Swarm : Immersive analytics for alerts classification based on periodic data », *Front. Virtual Real.*, t. 4, *April*, p. 1-17, 2023, issn : 26734192. doi : 10.3389/frvir.2023.1156656

6.1.3 Communications informelles

— Participations aux Rencontres Doctorales de la 33eme conférence internationale francophone sur l'Interaction Humains-Machines
— Présentation des travaux de thèse lors de la journée annuelle du groupe de travail Visualisation du groupement de recherche informatique géométrique et graphique, réalité virtuelle et visualisation (GdR IG-RV)

6.2 Autres communications

— Présentation de la Chaire Cyber CNI et de prototypes pendant les éditions 2021, 2022, et 2023, de la European Cyber Week
— Présentations des travaux de thèse aux partenaires industriels de la Chaire Cyber CNI
— Poster lors de l'événement de vulgarisation scientifique Science en Theiz

Annexe A

Annexes

A.1 Ensemble des papiers retenus pour l'étude de littérature

Nous avons sélectionné 50 articles pour notre revue de littérature. La page suivante présente, pour chaque document, ses objectifs dans le domaine de la cybersécurité.

	Citation	Tâches cyber							
Name		Host/Server	Internal/External	Port Activity	Attack Patterns	Routing Behavior	Vulnerability	User Monitoring	Accès Outils
A study on labeling network hostile behavior with intelligent interactive tools	[151]		✔✔✔	✔					oui
Alert characterization by non-expert users in a cybersecurity virtual environment : A usability study	[152]				✔✔✔				
BGP eye : a new visualization tool for real-time detection and analysis of BGP anomalies	[153]		✔			✔✔✔			
Bigfoot : A geo-based visualization methodology for detecting BGP threats	[154]					✔✔✔			
BUCEPHALUS : A BUsiness CEntric cybersecurity Platform for proActive anaLysis Using visual analyticS	[53]	✔			✔✔✔		✔✔✔		
Combining visual and automated data mining for near-real-time anomaly detection and analysis in BGP.	[155]					✔✔✔			oui
DAEDALUS-VIZ	[93]	✔	✔	✔					
Detecting malicious logins in enterprise networks using visualization	[156]				✔✔✔			✔✔✔	oui
Developing Visualisations to Enhance an Insider Threat Product : A Case Study	[157]						✔✔✔		
Ensemble visualization for cyber situation awareness of network security data	[64]			✔	✔✔✔				oui
ENTVis	[147]	✔	✔	✔	✔✔✔				
Existence plots : A low-resolution time series for port behavior analysis	[158]	✔	✔	✔✔✔					
Finding anomalies in time-series using visual correlation for interactive root cause analysis	[58]	✔	✔	✔	✔✔✔				oui
Flexible web visualization for alert-based network security analytics	[159]	✔	✔	✔	✔✔✔				oui
IDS RainStorm	[18]	✔	✔	✔	✔✔✔				oui
IMap : Visualizing network activity over internet maps	[160]	✔	✔	✔	✔✔✔				
Improving Interpretability for Cyber Vulnerability Assessment Using Focus and Context Visualizations	[161]						✔✔✔		
MAD : A visual analytics solution for Multi-step cyber Attacks Detection	[63]	✔✔✔	✔✔✔	✔✔✔	✔✔✔				oui
Mission-focused cyber situational understanding via graph analytics	[162]	✔✔✔			✔✔✔				oui
Network-wide intrusion detection supported by multivariate analysis and interactive visualization	[163]	✔	✔	✔	✔✔✔				oui
NflowVis	[164]			✔	✔✔✔				oui
Network intrusion visualization with niva, an intrusion detection visual and haptic analyzer	[96]	✔			✔✔✔				oui
NStreamAware : Real-time visual analytics for data streams to enhance situational awareness	[165]	✔	✔	✔	✔✔✔				
NV : Nessus vulnerability visualization for the web	[16]	✔					✔✔✔		oui
OCEANS	[60]	✔	✔	✔	✔✔✔				
Ocelot	[54]	✔			✔✔✔				oui
Percival	[57]	✔			✔✔✔		✔✔✔		
PortVis : A Tool for Port-Based Detection of Security Events	[166]			✔✔✔					
Preserving the big picture : visual network traffic analysis with TNV	[167]	✔	✔✔✔	✔					oui
Security in Process : Visually Supported Triage Analysis in Industrial Process Data	[121]								
Situ : Identifying and explaining suspicious behavior in networks	[51]	✔✔✔	✔✔✔	✔✔✔	✔✔✔				oui
SNAPS	[168]	✔		✔	✔✔✔			✔✔✔	oui
SnortView	[56]	✔		✔	✔✔✔				oui
SpiralView	[169]				✔✔✔		✔✔✔	✔✔✔	
SVision	[170]	✔	✔	✔	✔✔✔				
Tudumi : Information visualization system for monitoring and auditing computer logs	[99]	✔✔✔						✔✔✔	
Uncovering periodic network signals of cyber attacks	[26]	✔	✔	✔	✔✔✔				oui
Understanding the Context of Network Traffic Alerts	[171]	✔	✔	✔	✔✔✔				
VisTracer : A Visual Analytics Tool to Investigate Routing Anomalies in Traceroutes	[171]					✔✔✔			oui
Visual analysis of complex firewall configurations	[172]						✔✔✔		
Visual Firewall Log Analysis - At the Border Between Analytical and Appealing	[173]		✔✔✔	✔✔✔					
Visual firewall : Real-time network security monitor	[174]	✔	✔	✔	✔✔✔				oui
Visualization design for immediate high-level situational assessment	[175]	✔	✔	✔			✔✔✔		
Visualization of Host Behavior for Network Security	[176]	✔✔✔		✔					
Visualizing Automatically Detected Periodic Network Activity	[177]	✔	✔	✔	✔✔✔				
Visualizing cyber attacks using IP matrix	[59]	✔	✔	✔	✔✔✔				oui
Visualizing Network Security Events Using Compound Glyph	[55]	✔✔✔	✔	✔					
Visual correlation for situational awareness	[178]	✔		✔	✔✔✔				
Visualizing the insider threat : Challenges and tools for identifying malicious user activity	[123]	✔	✔	✔	✔✔✔			✔✔✔	
A 3D mixed reality visualization of network topology and activity results in better dyadic cyber team communication and cyber situational awareness	[90]	✔	✔	✔	✔✔✔				

A.2 Étude des systèmes de visualisation utilisés pour la cybersécurité

Étape 1 : sélection des alertes 30 papiers représentent cette étape.

Visualisation	Nombre de papiers
timeline	19
pixel scatterplot	4
glyphs	3
cyclique	3
tickets	3
force directed scatterplots	2
heatmap	2
raw	2
scatterplot	2
table	2
calendar	1
node/link graph	1
IDS	1
matrix	1
minimap	1
parallel	1
petri	1
radar	1
treemap	1

Étape 2 : identification des anomalies 47 papiers représentent cette étape.

Visualisation	Nombre de papiers
node/link graph	10
matrix	10
timeline	9
parallel	8
graph + glyphs	6
glyphs	4
pixel scatterplot	4
links	3
map	3
scatterplot	3
table	3
treemap	3
heatmap	2
Petri	2
cyclique	2
tree	2
dendogram	1
gantts	1
helix	1
hive	1
sunburst	1
text	1

Étape 3 : exploration des détails 30 papiers représentent cette étape.

Visualisation	Nombre de papiers
timeline	14
raw	15
table	7
node/link graph	3
matrix	3
glyphs	2
raw with colors	2
words clould	2
radar	1
treemap	1

Interactions Tous les papiers proposent des interactions avec le système de visualisation. La plupart des papiers proposent différents types d'interactions.

Interaction	Nombre de papiers
abstract	42
single	34
multiple	26
additional	23
direct	14
attributes	9
zoom	8
arrange views	4
arrange components	4
aggregate	3
gliding	3
brushing and linking	2
datapoints	2
highlighting	2
box	1
tresholds	1

Facettes et coordinations des vues 38 papiers utilise de multiples vues. Selon Munzner, si un système de visualisation propose différentes types de visualisations pour l'ensemble des données, il peut être catégorisé comme multiforme. Si un système de visualisation propose le même type de visualisation pour représenter un sous-ensemble du jeu de données, il peut être catégorisé comme overview+détail. Un système de visualisation peut combiner les deux attributs précèdents pour être multiforme avec overview + détail.

Facets	Nombre de papiers
different all	10
different all different subset	7
different subset	7
differant all same subset	6
same subset	3
same subset different subset different all	3
same subset different subset	2

Certains papiers peuvent proposer plusieurs techniques de coordinations en même temps.

Technique de coordination	Nombre de papiers
single and multi for additional	16
single for additional	16
juxtapose	13
highlight	11
superimpose	8
links	6
multi for subset	3
minimap	2

Actions prises par l'utilisateur .

Actions	Nombre de papiers
Annotate	9
Tune algorthim	7
Record	5
Take action	4
Import	2
Derive	1

Représentation de la baseline .

Visualisation	Nombre de papiers
timeline	23
Scatterplot	1
Calendar	1
cyclique	3
Animation	7
Graphs+Glyphs	5
Axis	10
Snapshot	10
Other	5
gantt	2

BIBLIOGRAPHIE

[1] C. I. S. Canada, « Strategic Early Warning for Criminal Intelligence », 2007. adresse : http://publications.gc.ca/collections/collection%7B%5C_%7D2013/sp-ps/PS64-107-2007-eng.pdf.

[2] T. Chandler, M. Cordeil, T. Czauderna et al., « Immersive Analytics », *2015 Big Data Vis. Anal. BDVA 2015*, p. 1-8, 2015. doi : 10.1109/BDVA.2015.7314296.

[3] T. Munzner, *Visualization Analysis & Design*. 2014, p. 1-397, isbn : 9781466508934. doi : 10.1201/b17511.

[4] D. Coss, « The CIA strikes back : Redefining confidentiality, integrity and availability in security », *J. Inf. Syst. Secur.*, p. 10, 2010, issn : 1551-0123. adresse : www.jissec.org.

[5] K. Y. Chai et M. F. Zolkipli, « Review on Confidentiality, Integrity and Availability in Information Security », *J. ICT Educ.*, t. 8, *2*, p. 34-42, 2021, issn : 22897844. doi : 10.37134/jictie.vol8.2.4.2021.

[6] M. R. Endsley, « Measurement of situation awareness in dynamic systems », *Hum. Factors*, t. 37, *1*, p. 65-84, 1995, issn : 00187208. doi : 10.1518/001872095779049499.

[7] C. Zhong, A. Alnusair, B. Sayger, A. Troxell et J. Yao, « AOH-Map : A Mind Mapping System for Supporting Collaborative Cyber Security Analysis », *Proc. - 2019 IEEE Conf. Cogn. Comput. Asp. Situat. Manag. CogSIMA 2019*, p. 74-80, 2019. doi : 10.1109/COGSIMA.2019.8724159.

[8] A. Kabil, T. Duval, N. Cuppens, G. Le Comte, Y. Halgand et C. Ponchel, « From Cyber Security Activities to Collaborative Virtual Environments Practices Through the 3D CyberCOP Platform », *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, t. 11281 LNCS, p. 272-287, 2018, issn : 16113349. doi : 10.1007/978-3-030-05171-6_14.

[9] M. Esteve, I. Pérez, C. E. Palau et al., « Cyber Common Operational Picture : A Tool for Cyber Hybrid Situational Awareness Improvement », *Cyber Def. Situat. Aware.*, p. 10, 2016. doi : 10.14339/STO-MP-IST-148-11-PDF. adresse : https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/Forms/Meeting%20Proceedings%20Document%20Set/docsethomepage.aspx?ID=42609%7B%5C&%7DFolderCTID=0x0120D5200078F9E87043356C409A0D30823AFA16F602008CF187B%5C&%7DList=7e2cc123-6186-4c30-8082-1ba0.

[10] S. C. Sundaramurthy, J. Case, T. Truong, L. Zomlot et M. Hoffmann, « A tale of three security operation centers », *Proc. ACM Conf. Comput. Commun. Secur.*, t. 2014-Novem, *November*, p. 43-50, 2014, issn : 15437221. doi : 10.1145/2663887.2663904.

[11] A. Torres, « Building a World-Class Security Operations Center : A Roadmap », *SANS Inst.*, *May*, p. 36, 2015.

[12] B. P. Hámornik et C. Krasznay, « Prerequisites of Virtual Teamwork in Security Operations Centers : Knowledge , Skills , Abilities », *Aarms*, t. 16, *3*, p. 73-92, 2017.

[13] A. Shah, R. Ganesan, S. Jajodia et H. Cam, « A methodology to measure and monitor level of operational effectiveness of a CSOC », *Int. J. Inf. Secur.*, t. 17, *2*, p. 121-134, 2018, issn : 16155270. doi : 10.1007/s10207-017-0365-1.

[14] P. Jacobs, A. Arnab et B. Irwin, « Classification of security operation centers », *2013 Inf. Secur. South Africa - Proc. ISSA 2013 Conf.*, 2013. doi : 10.1109/ISSA.2013.6641054.

[15] O. Akinrolabu, I. Agrafiotis et A. Erola, « The challenge of detecting sophisticated attacks : Insights from SOC analysts », *ACM Int. Conf. Proceeding Ser.*, 2018. doi : 10.1145/3230833.3233280.

[16] L. Harrison, R. Spahn, M. Iannacone, E. Downing et J. R. Goodall, « NV : Nessus vulnerability visualization for the web », *ACM Int. Conf. Proceeding Ser.*, p. 25-32, 2012. doi : 10.1145/2379690.2379694.

[17] G. A. Fink, C. L. North, A. Endert et S. Rose, « Visualizing cyber security : Usable workspaces », *6th Int. Work. Vis. Cyber Secur. 2009, VizSec 2009 - Proc.*, p. 45-56, 2009. doi : 10.1109/VIZSEC.2009.5375542.

[18] K. Abdullah et C. Lee, « IDS RainStorm : Visualizing IDS Alarms Workshop on Visualization for Computer Security », *Vizsec*, p. 1-10, 2005.

[19] H. I. Ansoff, « Managing Strategic Surprise by Response to Weak Signals », *Calif. Manage. Rev.*, t. 18, *2*, p. 21-33, 1975, issn : 21628564. doi : 10.2307/41164635.

[20] L. Methode, E. Sca et H. Lesca, « Lesca 2001 BARCELONE METHODE L », 2001.

[21] S. Sidhom et P. Lambert, « Information design for "weak signal" detection and processing in economic intelligence : A case study on health resources », *J. Intell. Stud. Bus.*, t. 1, *1*, p. 40-48, 2011, issn : 2001015X. doi : 10.37380/jisib.v1i1.13.

[22] S. Andrews, B. Brewster et T. Day, « Organised crime and social media : a system for detecting, corroborating and visualising weak signals of organised crime online », *Secur. Inform.*, t. 7, *1*, 2018, issn : 2190-8532. doi : 10.1186/s13388-018-0032-8. adresse : https://doi.org/10.1186/s13388-018-0032-8.

[23] G. A. Fink, C. L. North, A. Endert et S. Rose, « Visualizing cyber security : Usable workspaces », *6th Int. Work. Vis. Cyber Secur. 2009, VizSec 2009 - Proc.*, p. 45-56, 2009. doi : 10.1109/VIZSEC.2009.5375542.

[24] A. Valdes et S. Cheung, « Intrusion monitoring in process control systems », *Proc. 42nd Annu. Hawaii Int. Conf. Syst. Sci. HICSS*, p. 1-7, 2009. doi : 10.1109/HICSS.2009.273.

[25] R. Gove et L. Deason, « Visualizing Automatically Detected Periodic Network Activity », *2018 IEEE Symp. Vis. Cyber Secur. VizSec 2018*, p. 1-8, 2019. doi : 10.1109/VIZSEC.2018.8709177.

[26] N. Anh Huynh, W. Keong Ng, A. Ulmer et J. Kohlhammer, « Uncovering periodic network signals of cyber attacks », *2016 IEEE Symp. Vis. Cyber Secur. VizSec 2016*, 2016. doi : 10.1109/VIZSEC.2016.7739581.

[27] B. H. McCormick, T. A. DeFanti et M. D. Brown, « Definition of visualization », *ACM SIGGRAPH Comput. Graph.*, t. 21, *6*, p. 3-3, 1987, issn : 0097-8930. doi : 10.1145/41997.41998.

[28] S. K. Card, G. G. Robertson et J. D. Mackinlay, « The information visualizer, an information workspace », *Conf. Hum. Factors Comput. Syst. - Proc.*, p. 181-188, 1991. doi : 10.1145/108844.108874.

[29] C. Ware et P. Mitchell, « Visualizing graphs in three dimensions », *ACM Trans. Appl. Percept.*, t. 5, *1*, 2008, issn : 15443558. doi : 10.1145/1279640.1279642.

[30] J. J. Van Wijk et W. A. Nuij, « A model for smooth viewing and navigation of large 2D information spaces », *IEEE Trans. Vis. Comput. Graph.*, t. 10, *4*, p. 447-458, 2004, ISSN : 10772626. DOI : 10.1109/TVCG.2004.1.

[31] J. Mackinlay, « Automating the Design of Graphical Presentations of Relational Information », *ACM Trans. Graph.*, t. 5, *2*, p. 110-141, 1986, ISSN : 15577368. DOI : 10.1145/22949.22950.

[32] T. Munzner, « A nested model for visualization design and validation », *IEEE Trans. Vis. Comput. Graph.*, t. 15, *6*, p. 921-928, 2009, ISSN : 10772626. DOI : 10.1109/TVCG.2009.111.

[33] J. Bertin, « Semiology of graphics : diagrams, networks, maps », *Univ. Wisconsin Press*, 1983, ISSN : 0069-0805. DOI : 10.1080/00690805.1987.10438353.

[34] S. Stevens, « On the Theory of Scales of Measurement Author (s) : S . S . Stevens Published by : American Association for the Advancement of Science Stable URL : http ://www.jstor.org/stable/1671815 », *Science (80-.).*, t. 103, *2684*, p. 677-680, 1946.

[35] C. Ware, *Foundations for an Applied Science of Data Visualization*. 2013, p. 1-30, ISBN : 1558608192. DOI : 10.1016/b978-0-12-381464-7.00001-6.

[36] A. Cantu, « Proposition de modes de visualisation et d ' interaction innovants pour les grandes masses de données et / ou les données structurées complexes en prenant en compte les limitations perceptives des utilisateurs », 2018.

[37] R. J. Crouser, E. Fukuday et S. Sridhar, « Retrospective on a decade of research in visualization for cybersecurity », *2017 IEEE Int. Symp. Technol. Homel. Secur. HST 2017*, 2017. DOI : 10.1109/THS.2017.7943494.

[38] M. Brehmer et T. Munzner, « A multi-level typology of abstract visualization tasks », *IEEE Trans. Vis. Comput. Graph.*, t. 19, *12*, p. 2376-2385, 2013, ISSN : 10772626. DOI : 10.1109/TVCG.2013.124.

[39] A. Treisman et S. Gormican, « Feature Analysis in Early Vision : Evidence From Search Asymmetries », *Psychol. Rev.*, t. 95, *1*, p. 15-48, 1988, ISSN : 0033295X. DOI : 10.1037/0033-295X.95.1.15.

[40] C. Healey, « Visualization of multivariate data using preattentive processing », p. 1-122, 1992.

[41] O. Leslie Reiser et K. Koffka, « Principles of Gestalt Psychology. », *J. Educ. Psychol.*, t. 27, *4*, p. 310-313, 1935, ISSN : 0022-0663. DOI : 10.1037/h0052629.

[42] W. S. Cleveland et R. McGill, « Graphical perception : Theory, experimentation, and application to the development of graphical methods », *J. Am. Stat. Assoc.*, t. 79, *387*, p. 531-554, 1984, ISSN : 1537274X. DOI : 10.1080/01621459.1984.10478080.

[43] B. Shneiderman, « Eyes have it : a task by data type taxonomy for information visualizations », *IEEE Symp. Vis. Lang. Proc.*, p. 336-343, 1996, ISSN : 10492615. DOI : 10.1016/b978-155860915-0/50046-9.

[44] F. Van Ham et A. Perer, « "Search, show context, expand on demand" : Supporting large graph exploration with degree-of-interest », *IEEE Trans. Vis. Comput. Graph.*, t. 15, *6*, p. 953-960, 2009, ISSN : 10772626. DOI : 10.1109/TVCG.2009.108.

[45] S. K. Card, J. D. Mackinlay et B. Shneiderman, « Reviewing Readings in information Visualization : Using Vision to Think », *IEEE Multimed.*, t. 6, *4*, p. 93, 1999, ISSN : 1070986X. DOI : 10.1109/MMUL.1999.809241.

[46] A. M. Maceachren, « How maps work : representation, visualization and design », *How maps Work Represent. Vis. Des.*, *March*, 1995, ISSN : 1048-9053. DOI : 10.14714/cp24.757.

[47] J. Heer et M. Bostock, « Crowdsourcing graphical perception : Using mechanical Turk to assess visualization design », *Conf. Hum. Factors Comput. Syst. - Proc.*, t. 1, p. 203-212, 2010. DOI : 10.1145/1753326.1753357.

[48] L. Jiang, A. Jayatilaka, M. Nasim, M. Grobler, M. Zahedi et M. A. Babar, « Systematic Literature Review on Cyber Situational Awareness Visualizations », *IEEE Access*, t. 10, p. 57 525-57 554, 2022, ISSN : 21693536. DOI : 10.1109/ACCESS.2022.3178195. arXiv : 2112.10354.

[49] A. Komadina, Z. Mihajlovic et S. Gros, « Analysis of the Design Space for Cybersecurity Visualizations in VizSec », p. 1-11, 2022. DOI : 10.1109/vizsec56996.2022.9941422.

[50] H. Shiravi, A. Shiravi et A. A. Ghorbani, « A survey of visualization systems for network security », *IEEE Trans. Vis. Comput. Graph.*, t. 18, *8*, p. 1313-1329, 2012, ISSN : 10772626. DOI : 10.1109/TVCG.2011.144.

[51] J. R. Goodall, E. D. Ragan, C. A. Steed et al., « Situ : Identifying and Explaining Suspicious Behavior in Networks John », *IEEE Trans. Vis. Comput. Graph.*, t. 25, *1*, p. 204-214, 2019, issn : 19410506. doi : 10.1109/TVCG.2018.2865029.

[52] F. Zhou, W. Huang, Y. Zhao, Y. Shi, X. Liang et X. Fan, « ENTVis : A visual analytic tool for entropy-based network traffic anomaly detection », *IEEE Comput. Graph. Appl.*, t. 35, *6*, p. 42-50, 2015, issn : 02721716. doi : 10.1109/MCG.2015.97.

[53] M. Angelini, G. Blasilli, S. Bonomi et al., « BUCEPHALUS : A BUsiness CEntric cybersecurity Platform for proActive anaLysis Using visual analyticS », *Proc. - 2021 IEEE Symp. Vis. Cyber Secur. VizSec 2021*, p. 15-25, 2021. doi : 10.1109/VizSec53666.2021.00007.

[54] D. L. Arendt, R. Burtner, D. M. Best et al., « Ocelot : User-centered design of a decision support visualization for network quarantine », *2015 IEEE Symp. Vis. Cyber Secur. VizSec 2015*, p. 0–7, 2015. doi : 10.1109/VIZSEC.2015.7312763.

[55] J. Earlman et P. Rheingans, « Visualizing Network Security Events Using Compound Glyphs from a Service-Oriented Perspective », *Vizsec 2007*, 2007.

[56] H. Koike et K. Ohno, « SnortView : Visualization system of snort logs », *VizSEC/DMSEC '04 Proc. 2004 ACM Work. Vis. Data Min. Comput. Secur.*, p. 143-147, 2004.

[57] M. Angelini, N. Prigent et G. Santucci, « PERCIVAL : Proactive and reactive attack and response assessment for cyber incidents using visual analytics », *2015 IEEE Symp. Vis. Cyber Secur. VizSec 2015*, 2015. doi : 10.1109/VIZSEC.2015.7312764.

[58] F. Stoffel, F. Fischer et D. A. Keim, « Finding anomalies in time-series using visual correlation for interactive root cause analysis », *ACM Int. Conf. Proceeding Ser.*, p. 65-72, 2013. doi : 10.1145/2517957.2517966.

[59] H. Koike, K. Ohno et K. Koizumi, « Visualizing cyber attacks using IP matrix », *IEEE Work. Vis. Comput. Secur. 2005, VizSEC 05, Proc.*, p. 91-98, 2005. doi : 10.1109/VIZSEC.2005.1532070.

[60] S. Chen, C. Guo, X. Yuan, F. Merkle, H. Schaefer et T. Ertl, « OCEANS - Online collaborative explorative analysis on network security », *ACM Int. Conf. Proceeding Ser.*, t. 10-Novembe, p. 1-8, 2014. doi : 10.1145/2671491.2671493.

[61] A. Singh, L. Bradel, A. Endert, R. Kincaid, C. Andrews et C. North, « Supporting the cyber analytic process using visual history on large displays », *ACM Int. Conf. Proceeding Ser.*, 2011. doi : 10.1145/2016904.2016907.

[62] M. Schufrin, A. Ulmer, D. S. .-. 2. I. S. on . . . et U. 2018, « Towards Bridging the Gap Between Visual Cybersecurity Analytics and Non-Experts by Means of User Experience Design », *Vizsec.Org*, 2019. adresse : https://vizsec.org/files/2018/Schufrin%7B%5C_%7DPoster.pdf.

[63] M. Angelini, S. Bonomi, S. Lenti, G. Santucci et S. Taggi, « MAD : A visual analytics solution for Multi-step cyber Attacks Detection », *J. Comput. Lang.*, t. 52, *May 2017*, p. 10-24, 2019, issn : 25901184. doi : 10.1016/j.cola.2018.12.007. adresse : https://doi.org/10.1016/j.cola.2018.12.007.

[64] L. Hao, C. G. Healey et S. E. Hutchinson, « Ensemble visualization for cyber situation awareness of network security data », *2015 IEEE Symp. Vis. Cyber Secur. VizSec 2015*, 2015. doi : 10.1109/VIZSEC.2015.7312766.

[65] Paul Milgram et Fumio Kishino, « A Taxonomy of Mixed Reality Visual Displays », *IEICE Trans. Inf. Syst.*, *December 1994*, p. 1321-1329, 1996.

[66] P. P. Fuchs, *Le traité de la réalité virtuelle, t. 2*. Presses des MINES, 2006, p. 48.

[67] B. Ens, B. Bach, M. Cordeil et al., « Grand Challenges in Immersive Analytics », *Conf. Hum. Factors Comput. Syst. - Proc.*, 2021. doi : 10.1145/3411764.3446866.

[68] A. Prouzeau, Y. Wang, B. Ens, W. Willett et T. Dwyer, « Corsican Twin : Authoring in Situ Augmented Reality Visualisations in Virtual Reality », *ACM Int. Conf. Proceeding Ser.*, 2020. doi : 10.1145/3399715.3399743.

[69] M. Cordeil, B. Bach, A. Cunningham et al., « Embodied Axes : Tangible, Actuated Interaction for 3D Augmented Reality Data Spaces », *Conf. Hum. Factors Comput. Syst. - Proc.*, p. 1-12, 2020. doi : 10.1145/3313831.3376613.

[70] R. James, A. Bezerianos, O. Chapuis, M. Cordeil, T. Dwyer et A. Prouzeau, « Personal+Context navigation : Combining AR and shared displays in Network Path-following », *Proc. - Graph. Interface*, t. 2020-May, 2020, issn : 07135424. arXiv : 2005.10612.

[71] B. Ens, S. Goodwin, A. Prouzeau et al., « Uplift : A Tangible and Immersive Tabletop System for Casual Collaborative Visual Analytics », *IEEE Trans. Vis. Comput. Graph.*, t. 27, *2*, p. 1193-1203, 2021, issn : 19410506. doi : 10.1109/TVCG.2020.3030334.

[72] T. Chandler, M. Cordeil, B. Lee, B. Bach, T. Dwyer et K. Marriott, « Immersive Analytics », *IEEE Comput. Graph. Appl.*, t. 39, *3*, p. 16-18, 2019, issn : 15581756. doi : 10.1109/MCG.2019.2906513.

[73] J. Liu, A. Prouzeau, B. Ens et T. Dwyer, « Design and Evaluation of Interactive Small Multiples Data Visualisation in Immersive Spaces », *1*, p. 588-597, 2020. doi : 10.1109/vr46266.2020.00081.

[74] Y. Yang, T. Dwyer, K. Marriott, B. Jenny et S. Goodwin, « Tilt Map : Interactive Transitions between Choropleth Map, Prism Map and Bar Chart in Immersive Environments », *IEEE Trans. Vis. Comput. Graph.*, t. 27, *12*, p. 4507-4519, 2021, issn : 19410506. doi : 10.1109/TVCG.2020.3004137. arXiv : 2006.14120.

[75] K. A. Satriadi, B. Ens, M. Cordeil, T. Czauderna et B. Jenny, « Maps around Me : 3D Multiview Layouts in Immersive Spaces », *Proc. ACM Human-Computer Interact.*, t. 4, *ISS*, 2020, issn : 25730142. doi : 10.1145/3427329.

[76] K. A. Satriadi, B. Ens, M. Cordeil, B. Jenny, T. Czauderna et W. Willett, « Augmented reality map navigation with freehand gestures », *26th IEEE Conf. Virtual Real. 3D User Interfaces, VR 2019 - Proc.*, p. 593-603, 2019. doi : 10.1109/VR.2019.8798340.

[77] A. Ivanov, K. T. Danyluk et W. Willett, « Exploration & anthropomorphism in immersive unit visualizations », *Conf. Hum. Factors Comput. Syst. - Proc.*, t. 2018-April, p. 1-6, 2018. doi : 10.1145/3170427.3188544.

[78] C. Ware, D. Hui, G. Franck et C. Proceedings, « Visualizing Object Oriented Software in Three Dimensions University of New Brunswick CASCON ' 93 (IBM Centre for Advanced Studies) Visualizing Object Oriented Software in Three Dimensions University of New Brunswick Information visualization », t. 93, p. 612-620, 1993.

[79] R. Hackathorn et T. Margolis, « Immersive analytics : Building virtual data worlds for collaborative decision support », *2016 Work. Immersive Anal. IA 2016*, p. 44-47, 2017. doi : 10.1109/IMMERSIVE.2016.7932382.

[80] R. Skarbez, N. F. Polys, J. T. Ogle, C. North et D. A. Bowman, « Immersive Analytics : Theory and Research Agenda », *Front. Robot. AI*, t. 6, *September*, 2019, issn : 2296-9144. doi : 10.3389/frobt.2019.00082.

[81] A. Fonnet et Y. Prie, « Survey of Immersive Analytics », *IEEE Trans. Vis. Comput. Graph.*, t. 27, *3*, p. 2101-2122, 2021, issn : 19410506. doi : 10.1109/TVCG.2019.2929033.

[82] T. Dwyer, B. Bach, R. Dachselt, S. Carpendale, C. Collins et B. Lee, « Immersive analytics : Exploring future interaction and visualization technologies for data analytics », in *Proc. 2016 ACM Int. Conf. Interact. Surfaces Spaces Nat. Meets Interact. Surfaces, ISS 2016*, 2018, p. 529-533, isbn : 9781450342483. doi : 10.1145/2992154.2996365.

[83] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu et T. Isenberg, « The State of the Art of Spatial Interfaces for 3D Visualization », *Comput. Graph. Forum*, t. 40, *1*, p. 293-326, 2021, issn : 14678659. doi : 10.1111/cgf.14189.

[84] B. Bach, R. Sicat, J. Beyer, M. Cordeil et H. Pfister, « The Hologram in My Hand : How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality ? », *IEEE Trans. Vis. Comput. Graph.*, t. 24, *1*, p. 457-467, 2018, issn : 19410506. doi : 10.1109/TVCG.2017.2745941.

[85] J. P. McIntire et K. K. Liggett, « The (possible) utility of stereoscopic 3D displays for information visualization : The good, the bad, and the ugly », *2014 IEEE VIS Int. Work. 3DVis, 3DVis 2014, July 2015*, p. 1-9, 2015. doi : 10.1109/3DVis.2014.7160093.

[86] J. A. Wagner Filho, M. F. Rey, C. M. Freitas et L. Nedel, « Immersive Analytics of Dimensionally-Reduced Data Scatterplots », *25th IEEE Conf. Virtual Real. 3D User Interfaces, VR 2018 - Proc.*, p. 483-490, 2018.

[87] A. Fonnet, F. Melki, Y. Prié, F. Picarougne et G. Cliquet, « Immersive Data Exploration and Analysis », *Student Interact. Des. Res. Conf.*, p. 1-8, 2018. adresse : https://hal.archives-ouvertes.fr/hal-01798681/.

[88] E. H. Korkut et E. Surer, *Visualization in virtual reality : a systematic review*, 0123456789. Springer London, 2023, isbn : 0123456789. doi : 10.1007/s10055-023-00753-8. arXiv : 2203.07616. adresse : https://doi.org/10.1007/s10055-023-00753-8.

[89] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu et T. Isenberg, « The State of the Art of Spatial Interfaces for 3D Visualization », *Comput. Graph. Forum*, t. 40, *1*, p. 293-326, 2021, issn : 14678659. doi : 10.1111/cgf.14189.

[90] T. F. Ask, K. Kullman, S. Sütterlin, B. J. Knox, D. Engel et R. G. Lugo, « A 3D mixed reality visualization of network topology and activity results in better dyadic cyber team communication and cyber situational awareness », *Front. Big Data*, t. 6, 2023, issn : 2624909X. doi : 10.3389/fdata.2023.1042783.

[91] K. Kullman, M. Ryan et L. Trossbach, « VR/MR Supporting the Future of Defensive Cyber Operations », *IFAC-PapersOnLine*, t. 52, *19*, p. 181-186, 2019, issn : 24058963. doi : 10.1016/j.ifacol.2019.12.093. adresse : https://doi.org/10.1016/j.ifacol.2019.12.093.

[92] A. Kabil, T. Duval, N. Cuppens, G. Le Comte, Y. Halgand et C. Ponchel, « 3D CyberCOP : A collaborative platform for cybersecurity data analysis and training », *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, t. 11151 LNCS, p. 176-183, 2018, issn : 16113349. doi : 10.1007/978-3-030-00560-3_24.

[93] D. Inoue, M. Eto, K. Suzuki, M. Suzuki et K. Nakao, « DAEDALUS-VIZ : Novel real-time 3D visualization for darknet monitoring-based alert system », *ACM Int. Conf. Proceeding Ser.*, p. 72-79, 2012. doi : 10.1145/2379690.2379700.

[94] A. Cantu, T. Duval, O. Grisvard et G. Coppin, « HeloVis : A Helical Visualization for SIGINT Analysis Using 3D Immersion », *IEEE Pacific Vis. Symp.*, t. 2018-April, p. 175-179, 2018, issn : 21658773. doi : 10.1109/PacificVis.2018.00030.

[95] J. Gautier, P.-a. Davoine et C. Cunty, « Helical time representation to visualize return-periods of spatio-temporal events », 2017.

[96] C. Scott, K. Nyarko, T. Capers et J. Ladeji-Osias, « Network intrusion visualization with niva, an intrusion detection visual and haptic analyzer », *Inf.*

Vis., t. 2, *2*, p. 82-94, 2003, ISSN : 14738724. DOI : 10.1057/palgrave.ivs.9500044.

[97] L. LEICHTNAM, E. TOTEL, N. PRIGENT et L. ME, « STARLORD : Linked security data exploration in a 3D graph », *2017 IEEE Symp. Vis. Cyber Secur. VizSec 2017*, t. 2017-Octob, p. 1-4, 2017. DOI : 10.1109/VIZSEC.2017.8062203.

[98] C. WARE et G. FRANCK, « Viewing a graph in a virtual reality display is three times as good as a 2D diagram », *IEEE Symp. Vis. Lang. Proc.*, p. 182-183, 1994, ISSN : 10492615. DOI : 10.1109/vl.1994.363621.

[99] T. TAKADA et H. KOIKE, « Tudumi : Information visualization system for monitoring and auditing computer logs », *Proc. Int. Conf. Inf. Vis.*, t. 2002-Janua, *February*, p. 570-576, 2002, ISSN : 10939547. DOI : 10.1109/IV.2002.1028831.

[100] S. BUTSCHER, S. HUBENSCHMID, J. MÜLLER, J. FUCHS et H. REITERER, « Clusters, trends, and outliers : How Immersive technologies can facilitate the collaborative analysis of multidimensional data », *Conf. Hum. Factors Comput. Syst. - Proc.*, t. 2018-April, p. 1-12, 2018. DOI : 10.1145/3173574.3173664.

[101] P. KOBINA, T. DUVAL, L. BRISSON et A. DAVID, « Human-centered Evaluation of 3D Radial Layouts for Centrality Visualization », *Comput. Sci. Res. Notes*, t. 3201, *2022*, p. 86-94, 2022, ISSN : 24644625. DOI : 10.24132/CSRN.3201.9.

[102] R. F. ERBACHER, « Intrusion behavior detection through visualization », *Proc. IEEE Int. Conf. Syst. Man Cybern.*, t. 3, p. 2507-2513, 2003, ISSN : 08843627. DOI : 10.1109/icsmc.2003.1244260.

[103] F. CARROLL, A. CHAKOF et P. LEGG, « What makes for effective visualisation in cyber situational awareness for non-expert users ? », *2019 Int. Conf. Cyber Situational Awareness, Data Anal. Assessment, Cyber SA 2019*, p. 1-8, 2019. DOI : 10.1109/CyberSA.2019.8899440.

[104] B. BACH, E. PIETRIGA et J. D. FEKETE, « Visualizing dynamic networks with matrix cubes », in *Conf. Hum. Factors Comput. Syst. - Proc.*, Association for Computing Machinery, 2014, p. 877-886, ISBN : 9781450324731. DOI : 10.1145/2556288.2557010.

[105] B. Bach, P. Dragicevic, D. Archambault, C. Hurter et S. Carpendale, « A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes », *Comput. Graph. Forum*, t. 36, *6*, p. 36-61, 2017, issn : 14678659. doi : 10.1111/cgf.12804.

[106] E. Krokos, A. Rowden, K. Whitley et A. Varshney, « Visual Analytics for Root DNS Data », *2018 IEEE Symp. Vis. Cyber Secur. VizSec 2018*, p. 1-8, 2018. doi : 10.1109/VIZSEC.2018.8709205.

[107] L. Lisle, K. Davidson, E. J. Gitre, C. North et D. A. Bowman, « Sensemaking strategies with immersive space to think », *Proc. - 2021 IEEE Conf. Virtual Real. 3D User Interfaces, VR 2021*, p. 529-537, 2021. doi : 10.1109/VR50410.2021.00077.

[108] B. Lee, X. Hu, M. Cordeil, A. Prouzeau, B. Jenny et T. Dwyer, « Shared Surfaces and Spaces : Collaborative Data Visualisation in a Co-located Immersive Environment », t. 27, *2*, p. 1171-1181, 2021.

[109] L. Bradel, C. North, L. House et S. Leman, « Multi-model semantic interaction for text analytics », *2014 IEEE Conf. Vis. Anal. Sci. Technol. VAST 2014 - Proc.*, p. 163-172, 2015. doi : 10.1109/VAST.2014.7042492.

[110] A. Prouzeau, A. Lhuillier, B. Ens, D. Weiskopf et T. Dwyer, « Visual Link Routing in Immersive Visualisation », *28th Mod. Artif. Intell. Cogn. Sci. Conf. MAICS 2017*, p. 189-190, 2017. doi : 10.1145/1235.

[111] J. Liu, A. Prouzeau, B. Ens et T. Dwyer, « Design and Evaluation of Interactive Small Multiples Data Visualisation in Immersive Spaces », rapp. tech.

[112] A. Batch, A. Cunningham, M. Cordeil et al., « There Is No Spoon : Evaluating Performance, Space Use, and Presence with Expert Domain Users in Immersive Analytics », *IEEE Trans. Vis. Comput. Graph.*, t. 26, *1*, p. 536-546, 2020, issn : 19410506. doi : 10.1109/TVCG.2019.2934803.

[113] M. Cordeil, A. Cunningham, B. Bach et al., « IATK : An immersive analytics toolkit », *26th IEEE Conf. Virtual Real. 3D User Interfaces, VR 2019 - Proc.*, p. 200-209, 2019. doi : 10.1109/VR.2019.8797978.

[114] A. Hoppe, F. van de Camp et S. Rainer, « Enabling Interaction with Arbitrary 2D Applications in Virtual Environments Adrian », *Commun. Comput. Inf. Sci.*, t. 1225 CCIS, p. 30-36, 2020, ISSN : 18650937. DOI : 10.1007/978-3-030-50729-9_2.

[115] T. Mahmood, E. Butler, N. Davis, J. Huang et A. Lu, « Building Multiple Coordinated Spaces for Effective Immersive Analytics through Distributed Cognition », *2018 Int. Symp. Big Data Vis. Immersive Anal. BDVA 2018*, 2018. DOI : 10.1109/BDVA.2018.8533893.

[116] P. Reipschlager, T. Flemisch et R. Dachselt, « Personal augmented reality for information visualization on large interactive displays », *IEEE Trans. Vis. Comput. Graph.*, t. 27, *2*, p. 1182-1192, 2021, ISSN : 19410506. DOI : 10.1109/TVCG.2020.3030460. arXiv : 2009.03237.

[117] G. Perelman, E. Dubois, Probst, Alice et M. Serrano, « Visual Transitions around Tabletops in Mixed Reality : Study on a Visual Acquisition Task between Vertical Virtual Displays and Horizontal Tabletops », t. 6, *December*, 2022.

[118] Q. Ma et B. Millet, « Design Guidelines for Immersive Dashboards », *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, t. 65, *1*, p. 1524-1528, 2021, ISSN : 2169-5067. DOI : 10.1177/1071181321651177.

[119] A. Irlitti, R. T. Smith, S. V. Itzstein, M. Billinghurst et B. H. Thomas, « Challenges for Asynchronous Collaboration in Augmented Reality », *Adjun. Proc. 2016 IEEE Int. Symp. Mix. Augment. Reality, ISMAR-Adjunct 2016*, p. 31-35, 2017. DOI : 10.1109/ISMAR-Adjunct.2016.0032.

[120] C. Tominski et H. Schumann, « Enhanced Interactive Spiral Display », *Annu. Sigr. Conf. Spec. Theme Interact.*, *May*, p. 53-56, 2008. adresse : http://www.ep.liu.se/ecp/034/013/ecp083413.pdf.

[121] A.-P. Lohfink, S. D. D. Anton, H. D. Schotten, H. Leitte et C. Garth, « Security in Process : Visually Supported Triage Analysis in Industrial Process Data », *IEEE Trans. Vis. Comput. Graph.*, t. 26, *4*, p. 1638-1649, avr. 2020, ISSN : 1077-2626. DOI : 10.1109/TVCG.2020.2969007. arXiv : 1912.04865. adresse : https://ieeexplore.ieee.org/document/8968740/.

[122] S. Foresti, J. Agutter, Y. Livnat, S. Moon et R. Erbacher, « Visual Correlation of Network Alerts », *IEEE*, *8*, p. 1275-1279, 2006.

[123] P. A. Legg, « Visualizing the insider threat : Challenges and tools for identifying malicious user activity », *2015 IEEE Symp. Vis. Cyber Secur. VizSec 2015*, *October 2015*, 2015. doi : 10.1109/VIZSEC.2015.7312772.

[124] N. Anh Huynh, W. Keong Ng, A. Ulmer et J. Kohlhammer, « Uncovering periodic network signals of cyber attacks », *2016 IEEE Symp. Vis. Cyber Secur. VizSec 2016*, 2016. doi : 10.1109/VIZSEC.2016.7739581.

[125] P. Reipschlager, T. Flemisch et R. Dachselt, « Personal augmented reality for information visualization on large interactive displays », *IEEE Trans. Vis. Comput. Graph.*, t. 27, *2*, p. 1182-1192, 2021, issn : 19410506. doi : 10.1109/TVCG.2020.3030460. arXiv : 2009.03237.

[126] C. Tominski, P. Schulze-Wollgast et H. Schumann, « 3D information visualization for time dependent data on maps », *Proc. Int. Conf. Inf. Vis.*, t. 2005, p. 175-181, 2005, issn : 10939547. doi : 10.1109/IV.2005.3.

[127] M. Csikszentmihalyi, « Flow and the foundations of positive psychology : The collected works of Mihaly Csikszentmihalyi », *Flow Found. Posit. Psychol. Collect. Work. Mihaly Csikszentmihalyi*, p. 1-298, 2014. doi : 10.1007/978-94-017-9088-8.

[128] W. S. Cleveland et M. R, « Graphical Perception : Theory, Experimentation, and Application of Graphical Methods », *J. Am. Stat. Assoc.*, t. 17, *387*, p. 1-25, 2007. adresse : papers3://publication/uuid/048455A6-686D-43C6-834F-6C30A66D8F2F.

[129] K. Webga et A. Lu, « Discovery of rating fraud with real-time streaming visual analytics », *2015 IEEE Symp. Vis. Cyber Secur. VizSec 2015*, p. 1-8, 2015. doi : 10.1109/VIZSEC.2015.7312770.

[130] J. Brooke, « SUS : A 'Quick and Dirty' Usability Scale », *Usability Eval. Ind.*, *November 1995*, p. 207-212, 2020. doi : 10.1201/9781498710411-35.

[131] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt et J. Sun, « Anomaly detection for a water treatment system using unsupervised machine learning », *IEEE Int. Conf. Data Min. Work. ICDMW*, t. 2017-Novem, p. 1058-1065, 2017, issn : 23759259. doi : 10.1109/ICDMW.2017.149. arXiv : 1709.05342.

[132] S. D. Anton, A. P. Lohfink, C. Garth et H. D. Schotten, « Security in process : Detecting attacks in industrial process data », *ACM Int. Conf. Proceeding Ser.*, 2019. doi : 10.1145/3360664.3360669. arXiv : 1909.03730.

[133] C. C. M. Yeh, Y. Zhu, L. Ulanova et al., « Matrix profile I : All pairs similarity joins for time series : A unifying view that includes motifs, discords and shapelets », *Proc. - IEEE Int. Conf. Data Mining, ICDM*, p. 1317-1322, 2017, issn : 15504786. doi : 10.1109/ICDM.2016.89.

[134] T. Dwyer, K. Marriott, T. Isenberg et al., « Immersive Analytics : An Introduction To cite this version : HAL Id : hal-01907533 », 2018. doi : 10.1007/978-3-030-01388-2.

[135] J. A. Wagner Filho, C. M. Freitas et L. Nedel, « VirtualDesk : A Comfortable and Efficient Immersive Information Visualization Approach », *Comput. Graph. Forum*, t. 37, *3*, p. 415-426, 2018, issn : 14678659. doi : 10.1111/cgf.13430.

[136] M. Kraus, N. Weiler, D. Oelke, J. Kehrer, D. A. Keim et J. Fuchs, « The Impact of Immersion on Cluster Identification Tasks », *IEEE Trans. Vis. Comput. Graph.*, t. 26, *1*, p. 525-535, 2020, issn : 19410506. doi : 10.1109/TVCG.2019.2934395.

[137] A. Van Benschoten, A. Ouyang, F. Bischoff et T. Marrs, « MPA : a novel cross-language API for time series analysis », *J. Open Source Softw.*, t. 5, *49*, p. 2179, 2020. doi : 10.21105/joss.02179.

[138] S. G. Hart et L. E. Staveland, « Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research », *Adv. Psychol.*, t. 52, *C*, p. 139-183, 1988, issn : 01664115. doi : 10.1016/S0166-4115(08)62386-9.

[139] S. A. Jackson, R. C. Eklund et A. J. Martin, « The flow manual - The manual for the flow Scales », *Mind Gard.*, p. 4-85, 2010. adresse : http://www.mindgarden.com/products/flow.htm.

[140] G. Norman, « Likert scales, levels of measurement and the "laws" of statistics », *Adv. Heal. Sci. Educ.*, t. 15, *5*, p. 625-632, 2010, issn : 13824996. doi : 10.1007/s10459-010-9222-y.

[141] A. Bangor, P. Kortum et J. Miller, « Determining what individual SUS scores mean ; adding an adjective rating », *J. usability Stud.*, t. 4, *3*, p. 114-23, 2009.

[142] A. D. Prabaswari, C. Basumerda et B. W. Utomo, « The Mental Workload Analysis of Staff in Study Program of Private Educational Organization », *IOP Conf. Ser. Mater. Sci. Eng.*, t. 528, *1*, 2019, issn : 1757899X. doi : 10.1088/1757-899X/528/1/012018.

[143] M. Weber, M. Alexa et W. Müller, « Visualizing time-series on spirals », *Proc. IEEE Symp. Inf. Vis.*, p. 7-13, 2001. doi : 10.1109/infvis.2001.963273.

[144] X. Wang, L. Besançon, D. Rousseau, M. Sereno, M. Ammi et T. Isenberg, « Towards an Understanding of Augmented Reality Extensions for Existing 3D Data Analysis Tools », *Conf. Hum. Factors Comput. Syst. - Proc.*, p. 1-13, 2020. doi : 10.1145/3313831.3376657.

[145] A. H. Hoppe, L. Otto, F. van de Camp, R. Stiefelhagen et G. Unmüssig, « qVRty : Virtual keyboard with a haptic, real-world representation », *Commun. Comput. Inf. Sci.*, t. 851, p. 266-272, 2018, issn : 18650929. doi : 10.1007/978-3-319-92279-9_36.

[146] N. Moustafa et J. Slay, « UNSW-NB15 : A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) », *2015 Mil. Commun. Inf. Syst. Conf. MilCIS 2015 - Proc.*, p. 1-6, 2015. doi : 10.1109/MilCIS.2015.7348942.

[147] P. Fleck, A. Sousa Calepso, S. Hubenschmid, M. Sedlmair et D. Schmalstieg, « RagRug : A Toolkit for Situated Analytics », *IEEE Trans. Vis. Comput. Graph.*, 2022, issn : 19410506. doi : 10.1109/TVCG.2022.3157058.

[148] P. W. S. Butcher, N. W. John et P. D. Ritsos, « VRIA : A Web-based Framework for Creating Immersive Analytics Experiences », *IEEE Trans. Vis. Comput. Graph.*, p. 1-1, 2020, issn : 1077-2626. doi : 10.1109/tvcg.2020.2965109.

[149] N. Delcombel, A. Kabil, T. Duval et M.-O. Pahl, « CyberCopter : a 3D helical visualisation for periodic signals of cyber attacks », in *VR4SEC*, 2021, p. 1-5.

[150] N. Delcombel, T. Duval et M. O. Pahl, « Cybercopters Swarm : Immersive analytics for alerts classification based on periodic data », *Front. Virtual Real.*, t. 4, *April*, p. 1-17, 2023, issn : 26734192. doi : 10.3389/frvir.2023.1156656.

[151] J. L. Guerra, E. Veas et C. A. Catania, « A study on labeling network hostile behavior with intelligent interactive tools », *2019 IEEE Symp. Vis. Cyber Secur. VizSec 2019*, 2019. doi : 10.1109/VizSec48167.2019.9161489.

[152] A. Kabil, T. Duval et N. Cuppens, « Alert characterization by non-expert users in a cybersecurity virtual environment : A usability study », *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, t. 12242 LNCS, p. 82-101, 2020, issn : 16113349. doi : 10.1007/978-3-030-58465-8_6.

[153] S. T. Teoh, S. Ranjan, A. Nucci et C. N. Chuah, « BGP eye : A new visualization tool for real-time detection and analysis of BGP anomalies », *Proc. 3rd Int. Work. Vis. Comput. Secur. VizSEC'06. Co-located with 13th ACM Conf. Comput. Commun. Secur. CCS'06*, p. 81-90, 2006. doi : 10.1145/1179576.1179593.

[154] M. Syamkumar, R. Durairajan et P. Barford, « Bigfoot : A geo-based visualization methodology for detecting BGP threats », *2016 IEEE Symp. Vis. Cyber Secur. VizSec 2016*, 2016. doi : 10.1109/VIZSEC.2016.7739583.

[155] S. T. Teoh, K. Zhang, S. M. Tseng, K. L. Ma et S. F. Wu, « Combining visual and automated data mining for near-real-time anomaly detection and analysis in BGP », *VizSEC/DMSEC '04 Proc. 2004 ACM Work. Vis. Data Min. Comput. Secur.*, p. 35-44, 2004. doi : 10.1145/1029208.1029215.

[156] H. Siadati, B. Saket et N. Memon, « Detecting malicious logins in enterprise networks using visualization », *2016 IEEE Symp. Vis. Cyber Secur. VizSec 2016*, 2016. doi : 10.1109/VIZSEC.2016.7739582.

[157] M. Graham, R. Kukla, O. Mandrychenko, D. Hart et J. Kennedy, « Developing Visualisations to Enhance an Insider Threat Product : A Case Study », *Proc. - 2021 IEEE Symp. Vis. Cyber Secur. VizSec 2021*, p. 47-57, 2021. doi : 10.1109/VizSec53666.2021.00011.

[158] J. Janies, « Existence plots : A low-resolution time series for port behavior analysis », *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, t. 5210 LNCS, p. 161-168, 2008, issn : 03029743. doi : 10.1007/978-3-540-85933-8_16.

[159] L. Hao, C. G. Healey et S. E. Hutchinson, « Flexible web visualization for alert-based network security analytics », *ACM Int. Conf. Proceeding Ser.*, p. 33-40, 2013. doi : 10.1145/2517957.2517962.

[160] J. J. Fowler, M. Schneider, T. Johnson et al., « IMap : Visualizing network activity over internet maps », *ACM Int. Conf. Proceeding Ser.*, t. 10-Novembe, p. 80-87, 2014. doi : 10.1145/2671491.2671501.

[161] K. B. Alperin, A. B. Wollaber et S. R. Gomez, « Improving Interpretability for Cyber Vulnerability Assessment Using Focus and Context Visualizations », *2020 IEEE Symp. Vis. Cyber Secur. VizSec 2020*, p. 30-39, 2020. doi : 10.1109/VizSec51108.2020.00011.

[162] S. Noel, P. D. Rowe, S. Purdy, M. Limiero, T. Lu et W. Mathews, « Mission-focused cyber situational understanding via graph analytics », *Int. Conf. Cyber Conflict, CYCON*, t. 2018-May, p. 427-448, 2018, issn : 23255374. doi : 10.23919/CYCON.2018.8405029.

[163] R. Theron, R. Magan-Carrion, J. Camacho et G. M. I. Fernndez, « Network-wide intrusion detection supported by multivariate analysis and interactive visualization », *2017 IEEE Symp. Vis. Cyber Secur. VizSec 2017*, t. 2017-Octob, p. 1-8, 2017. doi : 10.1109/VIZSEC.2017.8062198.

[164] F. Fischer, F. Mansmann, D. A. Keim, S. Pietzko et M. Waldvogel, « Large-scale network monitoring for visual analysis of attacks », *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, t. 5210 LNCS, *January*, p. 111-118, 2008, issn : 03029743. doi : 10.1007/978-3-540-85933-8_11.

[165] F. Fischer et D. A. Keim, « NStreamAware : Real-time visual analytics for data streams to enhance situational awareness », *ACM Int. Conf. Proceeding Ser.*, t. 10-Novembe, p. 65-72, 2014. doi : 10.1145/2671491.2671495.

[166] J. McPherson, K.-L. Ma, P. Krystosk, T. Bartoletti et M. Christensen, « PortVis : A Tool for Port-Based Detection of Security Events », p. 73, 2004. doi : 10.1145/1029208.1029220.

[167] J. R. Goodall, W. G. Lutters, P. Rheingans et A. Komlodi, « Preserving the big picture : Visual network traffic analysis with TNV », *IEEE Work. Vis. Comput. Secur. 2005, VizSEC 05, Proc.*, p. 47-54, 2005. doi : 10.1109/VIZSEC.2005.1532065.

[168] B. C. Cappers et J. J. Van Wijk, « SNAPS : Semantic network traffic analysis through projection and selection », *2015 IEEE Symp. Vis. Cyber Secur. VizSec 2015*, 2015. doi : 10.1109/VIZSEC.2015.7312768.

[169] E. Bertini, P. Hertzog et D. Laianne, « SpiralView : Towards security policies assessment through visual correlation of network resources with evolution of alarms », *VAST IEEE Symp. Vis. Anal. Sci. Technol. 2007, Proc.*, p. 139-146, 2007. doi : 10.1109/VAST.2007.4389007.

[170] I. V. Onut et A. A. Ghorbani, « SVision : A novel visual network-anomaly identification technique », *Comput. Secur.*, t. 26, *3*, p. 201-212, 2007, issn : 01674048. doi : 10.1016/j.cose.2006.10.001.

[171] B. C. Cappers et J. J. Van Wijk, « Understanding the context of network traffic alerts », *2016 IEEE Symp. Vis. Cyber Secur. VizSec 2016*, 2016. doi : 10.1109/VIZSEC.2016.7739579.

[172] F. Mansmann, T. Göbel et W. Cheswick, « Visual analysis of complex firewall configurations », *ACM Int. Conf. Proceeding Ser.*, p. 1-8, 2012. doi : 10.1145/2379690.2379691.

[173] M. Schufrin, H. Lucke-Tieke et J. Kohlhammer, « Visual Firewall Log Analysis - At the Border Between Analytical and Appealing », p. 1-11, 2022. doi : 10.1109/vizsec56996.2022.9941462.

[174] C. P. Lee, J. Trost, N. Gibbs, R. Beyah et J. A. Copeland, « Visual firewall : Real-time network security monitor », *IEEE Work. Vis. Comput. Secur. 2005, VizSEC 05, Proc.*, p. 129-136, 2005. doi : 10.1109/VIZSEC.2005.1532075.

[175] R. F. Erbacher, « Visualization design for immediate high-level situational assessment », *ACM Int. Conf. Proceeding Ser.*, p. 17-24, 2012. doi : 10.1145/2379690.2379693.

[176] F. Mansman, L. Meier et D. A. Keim, « Visualization of host behavior for network security », *Math. Vis.*, p. 187-202, 2008, issn : 2197666X. doi : 10.1007/978-3-540-78243-8_13.

[177] R. Gove et L. Deason, « Visualizing Automatically Detected Periodic Network Activity », *2018 IEEE Symp. Vis. Cyber Secur. VizSec 2018*, 2018. doi : 10.1109/VIZSEC.2018.8709177.

[178] Y. Livnat, J. Agutter, S. Moon et S. Foresti, « Visual correlation for situational awareness », *Proc. - IEEE Symp. Inf. Vis. INFO VIS*, t. 1, p. 95-102, 2005, issn : 1522404X. doi : 10.1109/INFVIS.2005.1532134.

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : Cybersécurité en Réalité Virtuelle, améliorer le processus de détection d'intrusion, d'investigation et de décision via l'utilisation de techniques de visualisations 3D immersives

Mot clés : Cybersécurité, Réalité Virtuelle, Signaux Faible, Visualisation

Résumé : Dans cette thèse, nous avons examiné comment la réalité virtuelle pouvait contribuer à aider les opérateurs des centres d'opération cyber qui sont chargés de traiter un grand nombre d'alertes dans des délais restreints. Pour trier ces alertes, les opérateurs comparent le comportement du système surveillé avec son comportement nominal et doivent corréler des données nombreuses et variées. Les outils 2D dont ils disposent actuellement ne fournissent pas de visualisations efficaces. Celles-ci sont limitées par leurs difficultés à corréler des données entre plusieurs visualisations, et à représenter l'évolution du comportement d'un système au fil du temps. Nous avons donc créé un concept de visualisation 3D qui permet de dépasser ces limitations. Nous avons développé un prototype immersif basé sur ce concept que nous avons évalué par rapport à des visualisations 2D. Les résultats montrent une plus grande efficacité de notre prototype pour traiter des données complexes, tout en permettant engagement accru des utilisateurs au prix d'un effort physique supplémentaire. Nous avons ensuite élaboré un concept d'environnement immersif pour la cybersécurité qui permet d'utiliser des visualisations 3D tout en ayant accès à des outils classiques de cybersécurité. Nous avons démontré que sa réalisation est possible et avons commencé à créer un prototype limité mais fonctionnel.

Title: Virtual Reality for Cybersecurity, improve intrusion detection, investigation and decision-making processes through the use of immersive 3D visualization techniques

Keywords: Cybersecurity, Virtual Reality, Weak Signals, Visualization

Abstract: In this thesis, we examined how virtual reality could contribute to assisting operators in cyber operation centers who are responsible for processing a large number of alerts within tight deadlines. To sort through these alerts, operators compare the behavior of the monitored system with its nominal behavior and must correlate numerous and diverse data. The 2D tools they currently have do not provide effective visualizations. They are limited by their difficulty in correlating data between multiple visualizations, and in representing the evolution of a system's behavior over time. Therefore, we created a 3D visualization concept that overcomes these limitations. We developed an immersive prototype based on this concept, which we evaluated compared to 2D visualizations. The results show greater efficiency of our prototype in processing complex data, while allowing increased user engagement at the cost of additional physical effort. We thus devised a concept for an immersive cybersecurity environment that enables the use of 3D visualizations while having access to conventional cybersecurity tools. We then demonstrated that all the necessary software components for its realization are available and have begun combining them to create a limited but functional prototype.