Numéro d'ordre : **Université de Lille** **Année 2021**

THESE

présentée et soutenue publiquement le 30 Janvier 2021

En vue de l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ LILLE

Discipline : Informatique et applications

par

François BOUCHAUD

Analyse forensique des écosystèmes intelligents communicants de l'Internet des objets

Composition du jury

Président :	Pr. Nathalie ROLLAND	IRCICA, Université de Lille
Rapporteurs :	Pr. Issa TRAORÉ Dr. Anne-Cécile ORGERIE	Université de Victoria IRISA, Université de Rennes
Examinateur :	Pr. David BILLARD	Université des Sciences Appliquées de Genève
Invité :	Général d'armée Marc WATIN-AUGOUARD	Gendarmerie Nationale
Directeur de thèse :	Pr. Gilles GRIMAUD	CRIStAL, IRCICA, Université de Lille
Co-directeur de thèse :	Dr. Thomas VANTROYS	CRIStAL, IRCICA, Université de Lille

Centre de Recherche en Informatique, Signal et Automatique de Lille - Équipe 2XS

Remerciements

La thèse de doctorat soutenue est le fruit de trois années d'études au sein de l'équipe 2XS du groupe Systèmes Embarqués Adaptatifs et Sécurisés du Centre de Recherche en Informatique, Signal et Automatique de Lille (UMR 9189 du CNRS). Ce travail de recherche a été possible avec le soutien de nombreuses personnes que je souhaite vivement remercier.

Tout d'abord, je tiens à exprimer toute ma gratitude à Monsieur le Professeur Gilles GRIMAUD, directeur de thèse, pour avoir accepté d'assurer la direction de ce mémoire ainsi que son appréciation, sa disponibilité et ses conseils avisés lors de la réalisation. Il a adhéré très tôt au potentiel de ma recherche.

J'exprime ma reconnaissance à Monsieur le Docteur Thomas VANTROYS pour son soutien et sa compréhension tout au long de ce travail de thèse, de son implication dans les publications et de la relecture technique attentive des manuscrits. Á cette reconnaissance, j'associe également Monsieur le Docteur Alexandre BOÉ et le Chef d'Escadron Pierrick BURET pour leurs bons conseils, leurs soutiens techniques et leurs témoignages d'amitié à l'occasion de ces travaux.

J'exprime mon plus profond respect à tous mes supérieurs de gendarmerie pour leur confiance accordée à ma thèse de doctorat. Mes remerciements sont adressés particulièrement au Général Patrick TOURON, commandant du Pôle judiciaire de la Gendarmerie Nationale, pour avoir soutenu ce projet et en particulier auprès des plus hautes instances de la Gendarmerie Nationale dont la commission scientifique CEST. Je remercie toute la chaîne de commandement direct, le Colonel Franck MARESCAL, directeur de l'Institut de recherche criminelle de la Gendarmerie Nationale (IRCGN), le Colonel Fabrice BOUILLÉ, chef du Service central de renseignement criminel (SCRC), la Colonelle Fabienne LOPEZ, cheffe du Centre de lutte contre les criminalités numériques (C3N) et le Colonel Frédéric MONIN, directeur du Centre de formation des dirigeants de la gendarmerie (CFDG). Je n'oublierai jamais leurs contributions telles que je les comprends maintenant que les décisions des officiers soient nécessaires pour soutenir les projets auxquels nous croyons.

Je remercie chaleureusement les personnes qui ont contribué à la relecture et à la correction de cette thèse et de mes travaux.

Je remercie également les membres de CRIStAL qui m'ont accompagné tout au long de ma thèse, pour leurs conseils, la relecture et leurs aides précieuses. Je vous remercie notamment pour votre compréhension humaine et le haut niveau de compétence professionnelle.

Je suis reconnaissant envers tous ceux qui ont travaillé avec moi sur divers sujets, en particulier mes équipes à l'IRCGN et du C3N, officiers et sous-officiers. La diversité des profils et de leurs compétences constitue indéniablement la force des unités.

Un merci tout particulier pour le Maréchal des logis-chef Jean-Baptiste GRAFFIN et à l'ensemble des acteurs du CFDG qui m'ont toujours accordé un soutien apprécié pour m'aider dans mes différentes démarches administratives.

Je souhaite exprimer ma reconnaissance aux rapporteurs et aux examinateurs pour l'intérêt qu'ils portent à mon travail, pour avoir accepté de faire partie du jury de ma thèse de doctorat et pour m'avoir apporté leur soutien et leurs conseils.

Sur une note plus personnelle, je souhaite remercier, en particulier mes parents et tous mes proches. Sans eux, rien de tout cela n'aurait été possible.

Pontoise, 30 janvier 2021
François BOUCHAUD

Sommaire

Table des figures

Liste des tableaux

Première partie

Contexte et problématique

Chapitre 1

Introduction

Pour donner la vraie science du mouvement des oiseaux dans l'air, il est nécessaire d'établir d'abord la science des vents, laquelle explique les mouvements de l'eau et elle-même. Et cette science fera échelle pour venir à la connaissance des volatiles dans l'air et le vent.

Léonard de Vinci

La généralisation des technologies de l'information interactives dans le quotidien constitue, avec la prolifération des objets intelligents communicants, un enjeu d'actualité. Peu visibles, à l'exception des plus populaires d'entre eux comme par exemple les montres connectées, ils s'invitent, digitalisent et participent à notre quotidien. Tous les secteurs d'activité sont impactés par cette numérisation accélérée. L'Internet des objets (IdO, angl. *Internet of Things* (IoT)) suscite aujourd'hui autant de promesses d'opportunités économiques et sociétales que de questions voire d'inquiétudes, certaines de portée stratégique.

1.1 Contexte économique et sociétal

Avec un taux de croissance annuel moyen évalué à 11,3% entre 2020 et 2024 par *International Data Corporation* (IDC) [7], le marché de l'Internet des objets est en plein essor. Il représente un important vecteur de croissance économique. Ce secteur économique est estimé à 593 millions d'objets connectés portables en circulation dans le monde en 2019 [8]. D'ici 2023, chaque personne en France sera porteur, en moyenne, de 3,6 matériels connectés et passera plus de 18h chaque semaine sur Internet [9].

L'Internet des objets ouvre de nouvelles perspectives dans l'interconnexion des personnes et des biens avec intelligence. Cette structuration de l'informatique a pour volonté d'offrir une information correcte ou une action contextualisée au profit de destinataires ciblés, le tout à un moment déterminé. Cette solution polymorphe est basée sur des objets connectés (angl. *connected objects*), pouvant communiquer avec l'extérieur par une passerelle (angl. *gateway*), mais également entre eux de façon directe. Les points d'accès sont généralement locaux, à l'échelle d'un habitat ou d'une usine. Cette construction en partie décentralisée est mouvante. Elle constitue une mise en relation de différents micro-réseaux très hétérogènes. Un autre mode d'utilisation déconcentrée est le réseau de communication sans infrastructure, développé lors des manifestations à Hong-Kong en 2014 [10]. Cette structuration de l'espace numérique s'appuie sur une application exploitant un réseau Bluetooth. L'usage des technologies de communication, détourné de sa fonction primaire, se meut en nouvelles solutions ou services. Ainsi, l'architecture de l'Internet des objets est génératrice de réponses et de procédés industriels innovants.

Les objets de la vie quotidienne intègrent une multitude de capteurs et d'actionneurs. Ils génèrent alors de la donnée et la communiquent généralement à un faible débit. Cette caractéristique est principalement motivée par des contraintes énergétiques. La multiplication des objets déployés amène à un volume total d'informations échangées extrêmement élevé. Les données collectées constituent une nouvelle manne économique. Elles sont l'« or noir » du XXI$^{\text{ème}}$ siècle. Elles participent également à l'innovation dans la création de services personnalisés, en réponse aux besoins actuels et futurs. Ainsi, ce phénomène bouleverse les frontières traditionnelles. Il révolutionne les chaînes de valeur de l'entreprise et l'organisation territoriale. La limite entre le virtuel et le réel est en passe d'être abolie. Elle s'inscrit dans une logique de *continuum* cyberespace-espace physique [11].

Cette transformation accélérée du numérique génère des interrogations nouvelles sur la gestion des risques environnementaux, sanitaires ou économiques. L'utilisation croissante des objets connectés a de fortes implications en matière de sécurité et de confidentialité des données échangées. Les dispositifs détournés par malveillance de leurs usages premiers introduisent de nouveaux risques, des menaces d'atteintes numériques et économiques. Le cas récent d'un vol de données provenant du système d'information d'un casino aux États-Unis, passant par un thermomètre connecté d'aquarium, illustre ce phénomène grandissant [12]. L'environnement connecté se transforme en un vecteur criminel : de façon accessoire en facilitant la commission d'une infraction, de manière principale lorsqu'elle se rapporte au contenu ou bien en constituant son objet. Concrètement, cette singularité cybercriminelle se traduit de plusieurs manières : par des perturbations du fonctionnement nominal d'un dispositif connecté en l'empêchant de transmettre des données, par la prise de contrôle logique ou physique de l'environnement connecté en le détournement de son usage premier et/ou par un accès illégal

aux informations échangées ou stockées en portant atteinte aux données personnelles. Pour les entreprises, il s'agit de préserver la confiance des clients en garantissant la confidentialité des données personnelles, la sécurité des transactions ou la protection à l'égard des logiciels malveillants et des attaques informatiques. Á ces risques identifiés s'ajoutent la maîtrise et la supervision d'un parc d'objets en cohérence avec les habitudes et les usages de consommateurs de services. Par ailleurs, la mise en péril de la santé publique ou d'un écosystème est également réelle, s'agissant d'appareils dont l'utilisation a un lien direct avec la santé ou la sécurité. Ces menaces sont amplifiées par la diffusion et la massification accélérées de dispositifs composites dans un écosystème anarchique non régulé et non réglementé. Il souffre d'une insuffisance de vision globale en matière de sécurité (*security by design*). En effet, l'absence de standard de sécurité, l'hétérogénéité des protocoles et des technologies utilisés, le manque de bonnes pratiques en matière de conception, notamment dans le maintien en condition de sécurité, génèrent des risques élevés. Ainsi, les objets connectés sont vulnérables et sujets à des actions malveillantes pouvant mener à des menaces sérieuses pour une société de l'hyper-connectée.

La facilité de déploiement et d'utilisation des objets connectés devient également un nouveau défi pour les administrateurs des systèmes et des réseaux d'entreprise. Un déploiement anarchique de solutions par des salariés occasionne des perturbations des réseaux existants. La question de la détection et de l'identification des objets connectés devient donc un élément central pour garantir la sécurité des systèmes d'information.

1.2 Contexte et retombées stratégiques de l'IdO pour la défense et le judiciaire

La multiplication des capteurs entraîne une numérisation du réel. Le numérique se superpose et interagit avec le monde physique. Les objets connectés scrutent et interagissent avec notre quotidien. Ils génèrent une grande diversité d'informations : des données d'utilisateurs, de contexte, de système, des logs de fonctionnement, etc. Ainsi, tous les deux ans, la volumétrie des données générées par l'Internet des objets double la taille de l'univers numérique, il est estimé en 2020 à 44 000 milliards de giga-octets [13]. L'horizon de l'Internet des objets dans un écosystème numérique, désormais global, ouvre à d'importants enjeux et à des promesses d'opportunités pour le renseignement, la conduite opérationnelle, l'investigation judiciaire et la défense. Les objets connectés inscrits dans l'infrastructure de l'IdO créent une donnée longitudinale qui propose non seulement de pouvoir identifier et de fournir tous les éléments matériels nécessaires à la manifestation de la vérité judiciaire, mais aussi offre des champs à la prévention et à l'analyse des phénomènes. Le croisement des traces, sélectionnées avec intelligence, autorise des recoupements d'informations et des investigations inédites. La

presse internationale fait écho de plusieurs enquêtes criminelles incorporant des écosystèmes connectés. Des enquêteurs du Merseyside (Royaume-Uni) ont exploité les logs de fonctionnement et les données de géolocalisation d'une montre sportive *Garmin Forerunner 10* dans la résolution de l'affaire Paul Massey [14]. Les informations recueillies ont permis de reconstruire la chronologie des évènements survenus et de qualifier l'infraction en mettant en avant une préméditation du fait criminel. En Arkansas (États-Unis), l'appareil Amazon Echo a servi de témoin dans un meurtre en enregistrant les bruits ambiants [15]. Dans l'affaire Anthony Aiello en Californie (États-Unis), la correspondance entre les données Fitbit du bracelet connecté de la victime et les informations du système domotique a permis de confondre le meurtrier en contextualisant le crime [16]. L'objet connecté détourné de son usage premier offre des informations inédites pour l'investigation judiciaire. C'est notamment le cas des thermostats connectés disposant de facultés d'apprentissage. Couplés à l'écosystème de la maison, ces équipements sont en mesure de déclencher des actions automatiques tel l'allumage du chauffage lorsque le téléphone est reconnu dans un champ proche. Cette information est exploitable pour reconstituer la chronologie des événements révolus et déterminer des présences ou des déplacements dans un périmètre donné.

L'interception des données sensibles par l'exploitation des failles de sécurité des appareils connectés et des données systèmes s'avère primordiale pour l'investigation. L'authentification des agents communicants, l'exploitation et la contextualisation des données à des fins d'analyse, le croisement entre les informations recueillies, le contrôle de l'exposition des données nécessitent en effet une amélioration continue des techniques d'expertise. Pour l'investigation judiciaire, il s'agit d'anticiper l'apparition de nouvelles formes de criminalité, de gagner en agilité et en fiabilité, et dès lors répondre à la demande croissante d'expertise. L'objectif est également de limiter les usages impropres qui pourraient être faits des données. L'expert criminel en nouvelles technologies doit s'assurer du maintien de l'intégrité et de la qualité des données, de la collecte des éléments matériels pertinents à leur présentation devant une cour de justice. Or, la collecte des données se heurte à plusieurs difficultés : ces données numériques sont souvent dispersées et anonymisées, contraintes par des politiques propres de gestion. Leurs manipulations à des fins d'exploitation ou de conservation s'avèrent difficiles et sont sujettes à de potentielles altérations. Il est de plus essentiel de réaliser le chaînage des données pour obtenir une lecture de l'information lisible.

Ce constat est également applicable à la collecte d'information dans le cadre du renseignement. Un rapport publié le 1er février 2016 par le centre de recherche Berkman de l'université Harvard [17] estime que la quantité de données rassemblées par les objets connectés en fait l'une des pistes privilégiées pour que les agences de renseignement puissent contourner les protections mises en place sur les moyens de communication « classiques ». En 2016, James Clapper, directeur du renseignement national des États-Unis, a déclaré lors d'une audition

devant le Sénat américain « à l'avenir, les services de renseignements pourraient tirer parti de l'Internet des objets pour identifier, surveiller ou localiser des suspects, découvrir des indicateurs potentiels, ou obtenir des mots de passe » [18]. La proximité entre l'objet et l'humain pose des interrogations sur son potentiel destructeur dans le cadre d'un piratage d'une infrastructure connectée par un individu ou un groupe aux intentions hostiles. À ce phénomène, les objets ouvrent de nouvelles surfaces d'attaque pour intercepter des informations dans une logique de « guerre électronique ». L'utilisation de données personnelles ou de fonctionnement transitant sur les réseaux sans l'autorisation des propriétaires est la norme et complexifie la donne. Néanmoins, le risque est maitrisable par une connaissance fine des systèmes et de leur fonctionnement.

Enfin d'un point de vue opérationnel, les objets connectés sont des facteurs externes à prendre en considération dans un raisonnement tactique pour le succès d'une intervention. En effet, une manœuvre et son effet majeur peuvent être compromis par ce type de dispositifs et l'externalisation de la donnée. Par exemple, ils peuvent constituer une contrainte dans la recherche d'une discrétion ou d'un élément de surprise lors d'une progression. Détournés de leurs usages premiers, ils peuvent attenter à la vie des unités d'intervention. Il est donc primordial pour les unités d'identifier en amont les dispositifs afin d'adapter une réponse opérationnelle satisfaisante à la menace. Ainsi, les travaux de recherche développés dans cette thèse, abordés selon le spectre missionnel de l'investigation judiciaire dans l'Internet des objets, sont transposables au volet opérationnel et à l'intervention.

1.3 Objectif de la thèse et contributions

L'objet connecté est la face visible et locale de l'infrastructure de l'Internet des objets, porte d'entrée des investigations judiciaires. La recherche, l'identification et l'analyse de cet élément catalyseur sont primordiales pour comprendre l'architecture globale et obtenir une information pertinente au regard de l'enquête. L'enquêteur doit être en mesure d'associer à un phénomène criminel et sa donnée, un dispositif physique. Il doit ainsi comprendre le parcours de l'information dans l'architecture connectée, de son initialisation à son interception. Cette perception oriente les investigations et les actes techniques dans l'obtention de preuves pour le procès pénal.

L'identification des objets et de leur caractérisation technique sont la clef de voûte pour l'extraction et l'étude de l'information pertinente. Or, la diversité d'usage, de fonctionnement dans la remontée et la synchronisation de l'information, l'hétérogénéité des objets interdépendants rendent jusqu'à présent ce travail d'investigation très chronophage et fastidieux en l'absence d'une approche intégrale d'analyse. Parcellaire dans un objet, la donnée prend son

sens dans l'architecture globale. Comment appréhender avec intelligence les objets et leur environnement de connexion dans un contexte judiciaire ? Comment accéder à ce gisement d'information, eu égard aux nombreuses dépendances, aux liens cachés, à sa dispersion et à sa fragmentation dans l'infrastructure connectée ? Quelle crédibilité donner aux traces recueillies et reconstruites ? Sont-elles fiables et robustes pour l'enquête et donc présentables devant une cour de justice ?

Ces travaux de recherche proposent une méthodologie et des outils d'appréhension et d'analyse de l'environnement connecté, pour des enquêteurs du judiciaire. Elle s'appuie sur des pratiques issues de la criminalistique numérique classique, du réseau et de l'investigation dans l'informatique en nuage (*Cloud*). Les principales contributions de cette thèse sont :

- une méthodologie opérationnelle dans l'appréhension d'environnements connectés avec l'introduction du concept de « cartographie fréquentielle », d'outils de détection et de localisation par radiofréquence et des solutions dans la collecte et le placement sous-scellé des dispositifs connectés ;
- le développement d'une classification des objets connectés en fonction de critères liés à la donnée relevée, utilisée dans l'analyse et la reconstruction d'un développement criminel ;
- la proposition de solutions d'identification et de caractérisation d'un dispositif connecté au travers de techniques matérielles et logicielles.

1.4 Structure du document

Ce document de thèse s'articule autour de neuf chapitres, incluant un chapitre d'introduction et un chapitre de synthèse des travaux réalisés. Un exercice de scène de crime inspiré de plusieurs faits réels et contenant des objets connectés est décliné tout au long des chapitres. Il illustre et confronte les différentes méthodologies et outils proposés.

Le **chapitre 2** a pour objectif d'apporter au lecteur les éléments techniques et conceptuels nécessaires pour appréhender les travaux de thèse. Ce chapitre aborde et confronte les sciences forensiques au regard des grands principes fondateurs de la criminalistique numérique moderne et du concept de l'Internet des objets.

Le **chapitre 3** exprime les challenges et les problématiques dans l'appréhension et l'analyse d'un environnement contenant des dispositifs connectés, dans le cadre d'une enquête judiciaire. Il souligne l'importance d'identifier les sources de preuves pour l'IdO. Il développe le scénario d'illustration utilisé tout au long du manuscrit.

Le **chapitre 4** propose une classification des sources potentielles de preuves. Ce développement s'appuie sur des critères objectifs liés à la donnée et à l'efficience de la démarche. Cette partie introduit également une lecture de la scène de crime dans son appréhension.

Le **chapitre 5** détaille le processus de la détection et de la localisation des dispositifs connectés nécessaire à l'identification des sources de preuves. Il s'intéresse à la mesure de l'émission radioélectrique des équipements au travers de l'utilisation de la radio logicielle adaptée aux besoins métier.

Le **chapitre 6** développe le processus de la collecte des sources de preuves dans l'IdO. Il introduit une méthodologie de saisie des objets connectés et du réseau de communication ainsi que leurs conditionnements. Cette opération est réalisée afin de garantir l'intégrité des données présentes dans l'appareil connecté.

Le **chapitre 7** s'intéresse à l'analyse des informations recueillies par les dispositifs présents dans l'infrastructure IdO. Il développe une démarche basée sur l'étude temporelle, spatiale et contextuelle de la donnée, afin de reconstruire la chronologie des événements survenus.

Le **chapitre 8** rassemble un ensemble d'attributs facilitant l'étude d'un objet connecté. Ces éléments définissent une empreinte numérique caractérisant de la manière la plus fine et techniquement un dispositif. Cette démarche s'appuie sur ses éléments physiques (spectre électromagnétique, thermique, énergétique) et logicielles (taille des paquets réseau, latence, protocoles, etc.). Cette partie propose également d'approfondir l'identification et la classification des équipements au regard de l'émission acoustique des composants électroniques.

Le **chapitre 9** synthétise les travaux de thèse et la contribution scientifique apportée. Il ouvre à des perspectives de recherches pour les sciences forensiques en numérique.

1.5 Contributions scientifiques

Les travaux de cette thèse nous ont amenés à publier trois articles dans des conférences internationales [19, 20, 21], un article de journal [22] et deux articles de vulgarisation [23, 24].

Ces publications couvrent les différents aspects suivants :

— une partie de l'état de l'art sur l'Internet des objets et la criminalistique numérique (chapitre 2) exprimant le besoin d'approfondir la question de l'identification des dispositifs connectés (chapitre 3) [19, 22, 23] ;

— la classification des éléments composant l'infrastructure connectée au regard de la donnée et de paramètres liées à la criminalistique numérique (chapitre 4) [19, 22] ;

— la détection et la localisation des objets connectés présents localement sur une scène de crime (chapitre 5) [22, 20] ;

— la collecte des dispositifs locaux au regard de l'infrastructure connectée et des dépendances au réseau (chapitre 6) [21] ;
— l'analyse forensique des traces numériques (chapitre 7) [24].

Chapitre 2

État de l'art

Plusieurs thématiques doivent être précisées et revisitées car elles sont pertinentes et fondamentales dans la compréhension de la démarche et des orientations des travaux de la thèse. Ce chapitre aborde ainsi la question des sciences forensiques au regard des grands principes fondateurs de la criminalistique moderne et du concept d'Internet des objets. L'objectif est de corréler la démarche scientifique de l'étude des traces dans un environnement numérique étendu, polymorphe et connecté.

2.1 Sciences forensiques

La science forensique consiste en l'étude des traces initiées par une activité criminelle afin d'aider la justice à déterminer les causes et les circonstances d'un événement ou d'un phénomène. Elle s'appuie sur une démarche scientifique structurée et des méthodologies d'analyse rigoureuses.

2.1.1 Principes fondateurs de la criminalistique moderne

La science forensique s'est construite autour de plusieurs travaux fondateurs. Le postulat de Locard [25] stipule que « *tout contact laisse une trace* ». La rencontre entre deux éléments induit une interaction qui se traduit par un échange de matière et/ou une modification d'un environnement. La trace, que ce soit une marque, un signal ou un objet, est un signe apparent, pas nécessairement visible à l'œil nu. Elle est le vestige incomplet et imparfait d'une présence et/ou d'une action dans un endroit donné, à un moment daté [26]. L'absence de traces complète apporte une incertitude. Cependant, la convergence des informations réduit cette carence et soutient une cohérence de faits. Le postulat de Locard est complété par le principe de Kirk [27] en axant la réflexion sur l'identité, le processus d'identification et d'in-

dividualisation de la trace ainsi que de sa source. Il part du postulat que « *tout objet de notre univers est unique* ». Deux éléments, naturels ou artificiels, ont des caractéristiques propres qui les différencient et les individualisent. « *Une caractéristique majeure et insurmontable qui place la criminalistique à part des autres disciplines scientifiques est son intérêt pour l'individualisation. Les autres sciences se satisfont de la classification d'un objet dans une case de la taxonomie de la discipline considérée. La criminalistique cherche à relier l'objet à une source particulière* » [28]. L'objectif est de rechercher et de démontrer l'existence d'une origine unique liant des objets. Elle réside également à expliquer les éventuelles divergences. Ce point de convergence se présente sous différentes natures : une caractéristique biologique d'un individu, une matrice, un usinage, etc. Pour faciliter le rapprochement entre une trace et une source, des banques de données sont renseignées. Certaines tables contiennent des données d'individualisation comme pour l'*Acide Désoxyribonucléique* (ADN), les empreintes digitales ou la balistique, d'autres rassemblent des caractéristiques techniques offrant une catégorisation par groupe comme pour les stupéfiants, les explosifs, les peintures aérosols ou les traces d'outils et de semelles. Le rapprochement est également soumis à une incertitude. Les traces sont falsifiables ou/et altérables. Elles dépendent également du support et du milieu dans lesquels elles sont retrouvées. Un impact important réside dans le processus de collecte et de préservation lors des investigations. Ainsi, de nombreux facteurs de contamination et de dégradation sont à prendre en compte dans l'analyse. Les probabilités jouent un rôle central dans cette démarche en intégrant ce biais [29, 30, 31, 32, 33].

L'investigation criminelle s'appuie également sur le principe d'abduction développé par Peirce [34, 35]. Il s'agit d'établir les causes les plus vraisemblables à l'événement constaté et d'affirmer des hypothèses sur le fait criminel en étudiant le lien à une cause probable. Par exemple, dans le cadre d'une effraction d'une porte d'accès, la relation entre une trace d'outillage, le mode opératoire, les connaissances sur les caractéristiques des matériaux [36] et des objets usités s'appuie sur ce raisonnement. Le théorème Bayes-Laplace sur la probabilité des causes qui fut énoncé pour la première fois par le Révérend Bayes et dont Pierre-Simon Laplace su expliciter la profondeur, construit la connaissance sur la quantification des biais et des incertitudes. L'étude des éléments sériels est également pertinente dans le cadre de l'investigation criminelle. Elle se base sur le rapprochement et la mise en relation de propriétés communes sur des faits similaires. La source est souvent l'élément de convergence entre des faits. Elle est trahie par un mode opératoire singulier ou un usage d'objets semblables.

L'événement criminel est singulier, inhabituel, irréversible dans le temps et dans l'espace. La trace observée se détache de l'activité courante en toute objectivité et neutralité. L'objectif de cette science de l'analyse réside à décrire le cas dans son unicité. La matérialité de la trace offre de mesurer et de déterminer ses caractéristiques propres. Elle devient comparable à des données connues ou présentant des similarités. Cette information primaire, témoin d'un fait

passé, est mise en perspective avec ses semblables et un environnement. Elle devient un indice dans le processus judiciaire. Cet ensemble établit les circonstances et le fait criminel. Par leur structuration et leur appropriation par le juge, il se transforme en élément de preuve expertale dans le cadre du procès pénal.

2.1.2 Criminalistique numérique

La criminalistique numérique, appelé communément *digital forensic*, est une branche de la criminalistique moderne. Elle s'intéresse à la recherche de traces numériques contenues dans un média numérique ou une infrastructure informatique. Cette science est pluridisciplinaire, faisant appel à des connaissances en informatique, en électronique, en réseaux et en téléphonie. Elle regroupe ainsi plusieurs sous-branches ou sous-disciplines, telles que la criminalistique des réseaux de communication (*network monitoring*), de la mémoire [37, 38], des données et des fichiers supprimés et/ou morcelés (*carving*) [39, 40], des objets connectés [41, 42] et mobiles [43, 44, 45, 46, 47], de l'informatique nuagique (*cloud*) [48, 49, 50] et de la rétro-ingénierie avec notamment l'étude des *malwares* [51, 52, 53, 54, 44].

Elle s'accompagne d'une méthodologie d'étude de la scène numérique et des traces associées, en comprenant l'identification de l'incident et des preuves potentielles, la saisie, l'acquisition [55] et l'analyse [56, 57] des traces et des supports numériques, ainsi que la production d'une traçabilité associée. Cette discipline regroupe les analyses des supports en cours de fonctionnement, appelé *live forensic* et une approche dite *post-mortem* du matériel. L'acquisition idéale implique la capture d'une image de la mémoire volatile *Random Access Memory* (RAM) et la création d'une copie *bit-à-bit* exacte du support à étudier. Cette opération est réalisée si possible en utilisant un dispositif de blocage en écriture pour empêcher l'altération de l'original. Dans le cas de stockage en ligne de type *cloud computing*, l'acquisition est réalisée « en direct » selon une copie « logique » des données [55]. Après l'extraction des données, l'appel à une fonction de hachage est opéré afin garantir l'exactitude et l'intégrité des données copiées [58]. Elle constitue donc une référence lors de chaque nouvelle expertise.

2.2 Définition de l'Internet des Objets

L'origine de l'expression spécifique « Internet des objets » (IoT en anglais et IdO en français) est attribuée à Kevin Ashton en 1999, lors d'une présentation à Procter and Gamble. Elle apparaît également dans des travaux connexes dans le centre d'identification automatique du *Massachussetts Institute of Technology* (MIT).

2.2.1 Concept générique

L'Internet des objets désigne l'idée de connecter n'importe quel appareil physique à l'Internet. À ses débuts, elle fait référence à une mise en réseau d'objets équipés de puces radio-fréquence *Radio Frequency Identification* (RFID) pour procéder à leur suivi dans une chaîne d'approvisionnement. Cette notion s'est par la suite démocratisée et généralisée avec l'essor des réseaux sans-fil, du *cloud* et de la miniaturisation des systèmes embarqués. Parallèlement, les interprétations se diversifient. Pour certains, n'importe quel appareil connecté est une solution de l'Internet des objets ; pour d'autres, cette notion fait référence à l'analyse de données volumineuses et à la création de services en tant qu'aspect de l'IdO d'un produit. Une recherche sur ce concept révèle qu'aucune définition universelle et officielle n'existe à ce jour. Cette notion est un paradigme et non une chose tangible. En tant que paradigme, elle regroupe des technologies et des opérations qui doivent être comprises et adaptées au regard du domaine dans lequel elle est appliquée.

L'absence de définition ou de norme officielle, partagée par les différents acteurs et utilisateurs du domaine, laisse la place à la possibilité d'une appropriation. Les interprétations révèlent le biais inhérent de la personne qui l'utilise. Par exemple, la définition du *National Institute of Standards and Technology* (NIST) [59, 60] fait référence à l'Internet des objets en tant que système cyber-physique *Cyber-Physical System* (CPS), selon une approche industrielle de la chose. Elle a décortiqué le concept en liant les éléments d'infrastructure de l'IdO avec un comportement attendu.

La commission d'enrichissement de la langue française a adopté dans une publication au Journal Officiel au 11 janvier 2018 [61] la définition suivante : « *ensemble des objets connectés ainsi que des réseaux de télécommunication et des plates-formes de traitement des informations collectées qui leur sont associés* ». Cette définition est influencée par une approche réseau de la chose en se focalisant sur les aspects matériels de l'infrastructure de l'Internet des objets. De la même manière, la définition du département de la sécurité intérieure des États-Unis révèle un lien avec des infrastructures critiques en qualifiant l'IdO de « *connexion de systèmes et d'appareils à des fins essentiellement physiques (détection, chauffage / refroidissement, éclairage, commande de moteur, transport) à des réseaux d'information (y compris Internet) via des protocoles interopérables, souvent intégrés dans des systèmes embarqués* » [62].

Le groupe de travail « *Internet of Things – Global Standards Initiative* (IoT-GSI) » de l'*Union International des Télécommunication* (UIT) parle « *d'infrastructure mondiale de la société de l'information, qui permet de disposer de services évolués en interconnectant des objets (physiques ou virtuels) grâce aux technologies de l'information et de la communication interopérables existantes ou en évolution* » [63]. Cette définition insiste sur la création de valeur à travers de nouveaux services délivrés par l'infrastructure de communication. La

technologie demeure un vecteur physique pour le développement, l'agrégation et l'analyse de l'information. Cette approche est une référence pour plusieurs livres blancs donc celui de l'*Autorité de Régulation des Communications Electroniques et des Postes* (ARCEP), de Digital Security premier *Computer Emergency Response Team* (CERT) et par la commission des affaires économiques de l'Assemblée Nationale française [64]. Elle répond à une volonté de développer une stratégie de l'IdO en prenant en compte l'évolution technologique, le marché et les opportunités notamment en intégrant un volet cyber-sécurité. La définition officielle de l'IdO formulée par l'organisation internationale de normalisation *International Organization for Standardization* (ISO) et la commission électrotechnique internationale *International Electrotechnical Commission* (IEC) reprend cette vision centrée sur la création de valeur au travers de services en généralisant la notion d'infrastructure. Elle fait référence à une « *infrastructure d'entités, de personnes, de systèmes et de ressources d'information interconnectés avec des services qui traitent et réagissent aux informations du monde physique et du monde virtuel* ». Cette vision est complétée par un travail technique de normalisation. Par exemple, les ISO/IEC 20924:2018 et 30141:2018 proposent une architecture de référence normalisée, avec un vocabulaire commun, des schémas conceptuels réutilisables par les concepteurs et les développeurs d'applications.

La *Federal Trade Commission* (FTC), agence indépendante du gouvernement des États-Unis dans l'application du droit de la consommation et le contrôle des pratiques commerciales anticoncurrentielles, désigne l'Internet des objets en tant que dispositifs « *vendus ou utilisés par les consommateurs* ». Cette description renvoie vers la responsabilité en matière de protection des consommateurs. Au volet technologique et d'innovation, des aspects juridiques doivent être intégrés dans la compréhension de cette notion numérique.

La littérature scientifique définit ce concept comme « *un groupe d'infrastructure interconnectant les objets connectés et permettant leur gestion, l'exploration de données et l'accès aux données qu'ils génèrent* » [65]. Cette définition généraliste garde une certaine neutralité d'emploi. Elle intègre l'identification technologique, le contexte de fonctionnement, la création de valeur et les enjeux dans l'étude d'informations polymorphes. Elle fait donc référence à une notion d'architecture étendue, structurant un écosystème de dispositifs connectés, afin d'offrir de nouveaux services en créant de la valeur. Les travaux de recherche de la thèse se basent sur cette interprétation de la notion d'Internet des objets.

2.2.2 Architecture technique de l'Internet des objets

Maintenant que nous avons une définition établie de l'Internet des objets, il est important de décrire la structure du système technique. Une telle représentation de haut niveau est appelé une architecture de référence. Elle traduit une définition générale dans une définition

opérationnelle. L'Internet des objets repose sur la combinaison de nombreuses technologies matures en électronique, en informatique et en réseau (Figure 2.1). Il ne s'agit pas d'une technologie nouvelle mais d'un agrégat de solutions techniques offrant de nouveaux services dans la valorisation des données. La chaîne technique est particulièrement étendue. Elle regroupe plusieurs chaînes de valeur plus ou moins complexes. Elle constitue donc un ensemble complexe composé d'un écosystème « local » avec des objets connectés, des passerelles et des terminaux utilisateurs, mais également d'un environnement extérieur pour l'agrégation et la valorisation des données. La notion de « local » renvoie à l'environnement proche et physique contenant les dispositifs connectés. Il s'agit d'un espace clos comme un lieu d'habitation ou une construction industrielle mais également d'un espace ouvert structuré autour d'un référentiel comme par exemple avec l'usage de la montre connectée rattachée à son propriétaire. La zone de couverture associée est dépendante des protocoles utilisés par les objets pour communiquer.

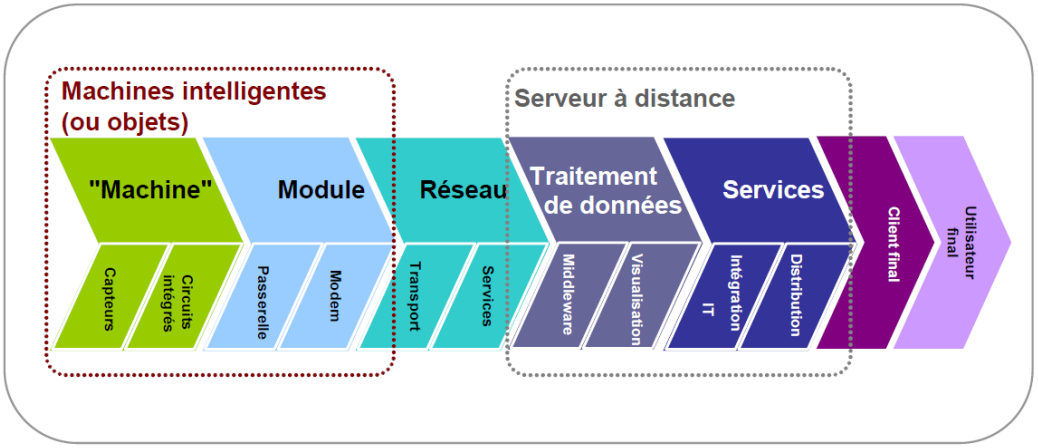


FIGURE 2.1 – Chaîne technique de l'IdO – Source : IDATE [1]

La mise en œuvre de l'IdO est généralement basée sur le traitement de données en temps réel avec des objets ou des périphériques sous-jacents, extrêmement limités en ressources. L'objet connecté est généralement un « petit » objet se révélant complexe en termes de fonctionnement et d'interactions. Bien qu'il existe une grande variété d'objets, il est possible de déconstruire le concept pour en faire ressortir les différents éléments constitutifs (cf. Figure 2.2). L'objet est généralement autonome en énergie et de faible puissance. Pour répondre à un usage, il possède des capteurs et/ou des actionneurs. Il traite des données la plupart du temps via un microcontrôleur. Ainsi, sa capacité de traitement ou de stockage est limitée par ces contraintes physiques [66]. Il reçoit ou envoie les informations collectées au travers d'un réseau de communication. La décision de la quantité de données à transmettre ou de la transmission des données traitées ou non traitées est également affectée par les ressources limitées disponibles. Il est donc intrinsèquement complexe, en liant le matériel et le logiciel

embarqué. Ces différents objets sont interconnectés au travers de divers médiums. Localement, la liaison entre les objets et la passerelle utilise généralement un lien radiofréquence (Bluetooth, *Wireless Fidelity* (Wi-Fi), *Light Fidelity* (Li-Fi), ZigBee, ZWave, Sigfox, *Long Range* (LoRa), etc.). Les couches physiques et les protocoles de niveau supérieur employés sont très variés. En effet, il existe une très forte hétérogénéité des protocoles de communication. De plus, certains objets non manufacturés, dits « fait maison », utilisent des protocoles de communication non standards. Il faut également noter que les objets peuvent utiliser des protocoles *ad hoc*. Les passerelles et les terminaux utilisateurs sont plus classiques dans leur fonctionnement et leur architecture. Leurs systèmes d'exploitation permettent de faire une abstraction importante entre le logiciel et le matériel sur lesquels ils s'exécutent.

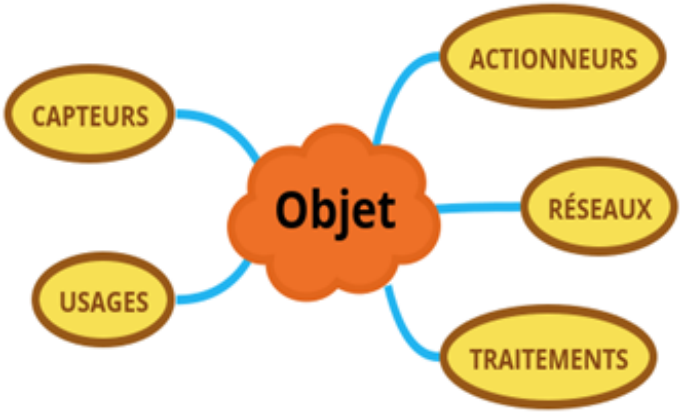


FIGURE 2.2 – Éléments constitutifs d'un objet connecté – Source : [2]

Pour communiquer avec le monde extérieur, les objets connectés utilisent des passerelles. Ces dispositifs de communication autorisent de réaliser une conversion ou une adaptation des protocoles. Les passerelles se trouvent soit localement auprès de l'équipement connecté comme dans le cas d'une box Wi-Fi ou bien soit elles sont gérées par un opérateur extérieur comme dans le cas du protocole SigFox.

L'utilisateur d'un portail web (ordinateur personnel) ou d'application mobile (tablette graphique ou *smartphone*) interagit avec l'objet, directement ou via une passerelle, en vue de réaliser une action et/ou de recueillir les données locales. À partir de son *Interfaces Homme-Machine* (IHM), il consulte les données stockées en ligne. La liaison est soit filaire dans le cas d'un terminal fixe, soit par radiofréquence pour les solutions mobiles. Elle s'appuie sur les protocoles ainsi que les couches physiques traditionnelles des réseaux (principalement TCP/IP avec *Transmission Control Protocol* (TCP) et *Internet Protocol* (IP)). La figure 2.3 reprend les différents niveaux de l'Internet des objets.

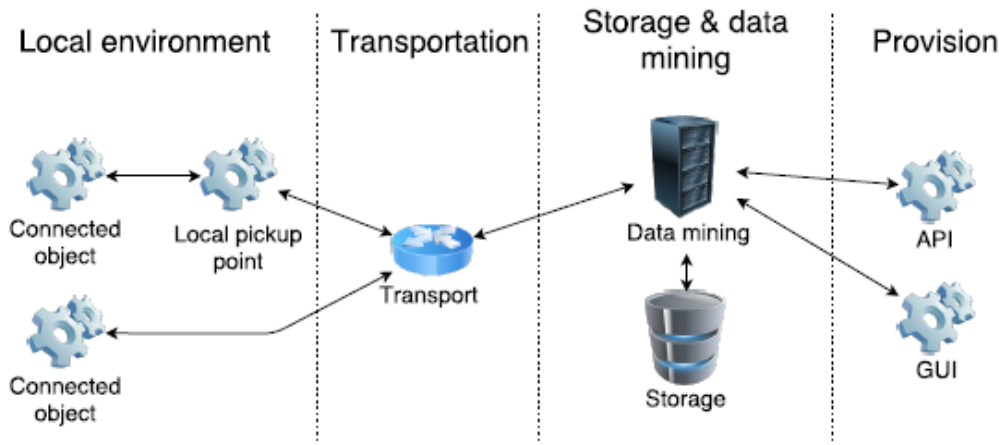


FIGURE 2.3 – Vue schématique de l'infrastructure de l'Internet des Objets – Source : [3]

Analyser un objet connecté consiste à comprendre la façon dont l'objet se connecte au réseau, comment il est contrôlé par l'application mobile associée et les interactions avec le serveur distant.

2.3 Sciences forensiques dans l'Internet des objets

Le fossé entre le monde physique et le monde numérique se réduit avec la massification du nombre d'appareils connectés à Internet, conséquence directe de l'apparition de l'Internet des objets. Les objets interfèrent avec l'environnement immédiat en le scrutant et en l'interrogeant. Ils constituent des réceptacles de l'information et génèrent un flux massif de données dans le système d'information. Face à ce constat, la criminalistique de l'Internet des objets prend tout son sens et devient une opportunité inédite dans la résolution de faits criminels.

2.3.1 Potentiel criminalistique de l'IdO

L'Internet des objets est une convergence de l'Internet et des réseaux de capteurs. Il demeure donc une extension des appareils numériques traditionnels comme les ordinateurs, les *smartphones* ou bien la domotique dans lesquels l'interconnexion et l'interaction sont au cœur du fonctionnement. Ainsi, l'Internet des objets fournit une infrastructure commune pour les entités du monde réel, vivantes ou non, qui créent et partagent des données sur Internet.

Les données recueillies dans cette infrastructure plurielle constituent une mine d'information en particulier dans la recherche de la vérité, en tant que témoin ou acteur privilégié d'un

fait passé. Elles apportent une aide dans la reconstruction et la transcription des événements survenus. Comparée aux techniques d'investigation numérique standard, l'étude de l'Internet des objets relève de multiples défis liés à la polyvalence et à la complexité des dispositifs trouvés localement. Ainsi, les objets connectés sont hétérogènes dans leurs natures et leurs fonctionnements. Ils disposent de logiciels et de briques matérielles propriétaires. Certains objets sont en mesure de se connecter directement à Internet, d'autres sont dépendants de passerelles dédiées. L'Internet des objets regroupe ainsi une structure physique locale ramifiée vers l'extérieur, au travers du *cloud* et de l'Internet. La donnée, en parcourant l'ensemble de ces espaces, n'est plus liée à un support réel et unique. Le défi est amplifié par la politique de gestion de la donnée, que ce soit sur la question du stockage et du traitement mais également que ce soit sur la remontée et sur la synchronisation de l'information dans le réseau. Elle se retrouve alors fragmentée et dispersée au sein de cet environnement pluriel. Elle est en perpétuelle évolution. Parcellaire dans un objet ou un écosystème, elle devient un tout dans l'architecture globale. Pour l'appréhender, l'enquêteur s'appuie sur les pratiques et outils alliant toutes les compétences de la criminalistique numérique (*digital forensic*).

2.3.2 Concept de criminalistique de l'IdO

La criminalistique de l'Internet des objets (*IoT forensic*) constitue donc une branche de la criminalistique numérique. Elle traite de l'étude des traces initiées par une activité criminelle dans un environnement numérique et connecté. Elle s'intéresse aux objets connectés présents sur la scène d'infraction, aux interfaces avec le réseau et les utilisateurs, ainsi qu'aux données stockées en ligne. Elle regroupe plusieurs domaines complémentaires de l'investigation criminelle : l'analyse des données dans le *cloud*, l'étude des réseaux et de l'électronique connectée. Elle apporte également des nouveautés dans l'appréhension et l'analyse de cet environnement multiforme. En-effet, l'Internet des objets renvoie à une vision du tout connecté, composée de fortes dépendances. Chaque thématique ne peut être pensée d'une manière exclusive.

Bien que le domaine de l'Internet des objets en matière d'investigation numérique soit relativement nouveau, il existe de nombreux travaux scientifiques intéressants. Dans Hegarty et al. [67], Zulkipli et al. [68] et Servida et al. [69], les auteurs discutent des défis fondamentaux que pose la criminalistique numérique pour l'Internet des objets. Dans Oriwoh et al.[70], ils se demandent comment la criminalistique numérique dans un environnement connecté s'éloigne de l'approche traditionnelle. Pour faciliter l'appréhension de cette structure, ils définissent également un découpage par zone. Dans Perumal et al.[4], ils donnent des instructions sur la manière de mener à bien l'investigation à partir de cette approche. Dans Kebande et al. [71], ils développent un cadre générique d'investigation composé de trois modules : le processus proactif, l'investigation de l'IdO, le processus réactif et les processus concurrents englobant les

trois modules susmentionnés. Le processus proactif suggère que la police scientifique numérique soit prête à mettre en place une planification et des processus pour traiter les incidents. Dans Copos et al. [72] et Zawoad et al. [73], les auteurs proposent une interface de collecte des données dans l'infrastructure connectée, sur les plates-formes en ligne ou directement dans le réseau local. Dans Rahman et al. [74], ils tentent d'identifier les sources de preuves en se basant sur des scénarios d'attaque. La littérature scientifique est riche de solutions dans la collecte et l'analyse des données présentes dans l'infrastructure connectée. Cependant, elle part du postulat que l'objet et son environnement sont identifiés et connus techniquement par l'enquêteur. Or, la scène à étudier demeure une inconnue dans chaque investigation. Ainsi, les processus de recherche et d'appréhension demeurent peu étudiés. Cette étape initiale constitue le fondement de la criminalistique numérique visant à trouver des preuves cohérentes au regard des éléments d'enquête.

Ce domaine d'étude est cependant confronté à des limites existentielles, en particulier sur sa partie *cloud* [75, 76, 77, 78, 79, 80]. La question de la localisation des données est récurrente [81, 82]. Á des contraintes techniques s'ajoutent des difficultés liées aux juridictions propriétaires ou géographiquement différentes et à l'absence d'accord avec les opérateurs de plates-formes [70]. La coopération internationale entre les acteurs du monde numérique est nécessaire pour le succès des opérations d'investigation [83]. En fonction des pays, la communication des informations stockées sur les serveurs est une obligation légale pour les prestataires de service, comme avec le *Cloud Act*[1], que les données soient situées aux États-Unis ou à l'étranger. Le contrôle et l'accès à la donnée est limitée dans la saisie de l'équipement numérique. Seule l'intervention de l'exploitant de la plate-forme à la suite d'une réquisition judiciaire est à même de répondre au besoin de la collecte de preuves. En outre, les services dans le *cloud* sont basés sur des machines virtuelles comme serveurs. Les données volatiles telles que les entrées de registre ou les fichiers Internet temporaires de ces serveurs sont effacées en l'absence de synchronisation avec les périphériques de stockage. Plusieurs travaux scientifiques proposent des modèles ou des solutions pour résoudre les problèmes inhérents à la préservation des preuves numériques provenant du *cloud* [84, 85, 86]. Parallèlement aux réquisitions réalisées auprès des opérateurs de plates-formes, l'enquêteur peut également exploiter l'accès au *cloud* au travers d'un client. Cette approche consiste à acquérir et à analyser les données enregistrées localement par des applications ou des navigateurs web en relation avec l'utilisation de services dans les nuages. Cette action doit cependant respecter le strict cadre légal de la perquisition en ligne. Dans Roussev et al. [87], les auteurs présentent des acquisitions de données en exploitant les arte-

1. Le *Cloud Act* (*Clarifying Lawful Overseas Use of Data Act*) est une loi fédérale américaine promulguée le 23 mars 2018. Elle modifie principalement le chapitre 121 du Titre 18 du *United States Code*, dénommé *Stored Communications Act*.

facts natifs de *Dropbox*, *OneDrive*, *Microsoft*, *Google Drive* via des *Application Programming Interface* (API) prises en charge par les services de *cloud*.

2.4 En quelques mots : criminalistique de l'Internet des objets

L'Internet des objets constitue une opportunité inédite pour la criminalistique numérique, science dans la recherche et l'étude des traces numériques. Cependant, elle est confrontée à des écosystèmes hétérogènes, structurés autour d'une architecture ouverte où la politique de la gestion de la donnée se réinvente pour chaque environnement. Plurielle dans sa définition et sa démarche, elle doit s'adapter à ces nouvelles contraintes d'usage par l'identification des objets physiques, par l'étude de l'infrastructure et de ses dépendances, en passant par des plates-formes *cloud* décentralisées.

Chapitre 3

Problématique

La criminalistique de l'Internet des objets est un défi et une opportunité pour les enquêteurs judiciaires. Cet environnement connecté est pluriel. Il regroupe des dispositifs hétérogènes : objets connectés et passerelles de communication. À cet assortiment, se greffe un réseau polymorphe liant les différents écosystèmes en local, tout en développant une ramification extérieure vers des plates-formes de traitement en ligne. Le tout est interconnecté à des interfaces de commandes et/ou à d'autres environnements ouverts. Cette organisation informatique est gouvernée par des politiques de gestion de la donnée non standardisées, tant au niveau du stockage que de la remontée de l'information dans le réseau. Ainsi, elle offre un gisement conséquent de données, traduisant les habitudes des utilisateurs et les états de fonctionnement des systèmes informatiques.

3.1 Données au cœur de l'enquête judiciaire

L'Internet des objets contribue à la numérisation du quotidien et des phénomènes éphémères. Il propose de nouvelles sources de preuves pour l'investigation judiciaire. Ces éléments délivrent des données exploitables sous de nombreux formats, plus ou moins standardisés.

3.1.1 Nouvelles perspectives offertes par l'Internet des objets

Ce domaine du tout connecté ouvre de nouvelles perspectives à la criminalistique numérique moderne, en particulier sur la question des sources [88]. Celles-ci diffèrent par leur nature, leur nombre, leur format et les protocoles utilisés. Classiquement, les matériels étudiés sont des ordinateurs, des appareils mobiles, des passerelles, des équipements de stockages ou des serveurs. En ce qui concerne l'Internet des objets, les preuves sont extraites des appareils ménagers, de systèmes domotiques, des équipements médicaux pour le vivant - homme

ou animal -, des voitures, des lecteurs RFID, etc. [89, 90]. Les données varient également en fonction des interactions avec l'environnement et des services fournis.

Face à cette démultiplication des sources de preuves, l'enquêteur a besoin de prioriser son action. Une classification efficace de l'environnement doit être élaborée afin de déterminer les informations pertinentes pour l'enquête. Existe-t-il des critères objectifs répondant à ce besoin de hiérarchisation ? Est-ce que cette catégorisation des sources est généralisable à tous les environnements connectés ? Quelles sont ces limites et ces contraintes ?

3.1.2 Richesse de la preuve par la pluralité des données échangées

L'environnement est composé des objets connectés, de leurs passerelles, de l'infrastructure de communication, des plates-formes et d'interfaces. Chaque élément retourne une multitude d'information (cf. Figure 3.1). La donnée peut être liée à l'événement et/ou à son contexte de réalisation, mais également au système de communication. Elle se présente sous une multitude de formats en fonction des équipements présents et de leurs rôles dans la chaîne de collecte, de transmission et de traitement de l'information.

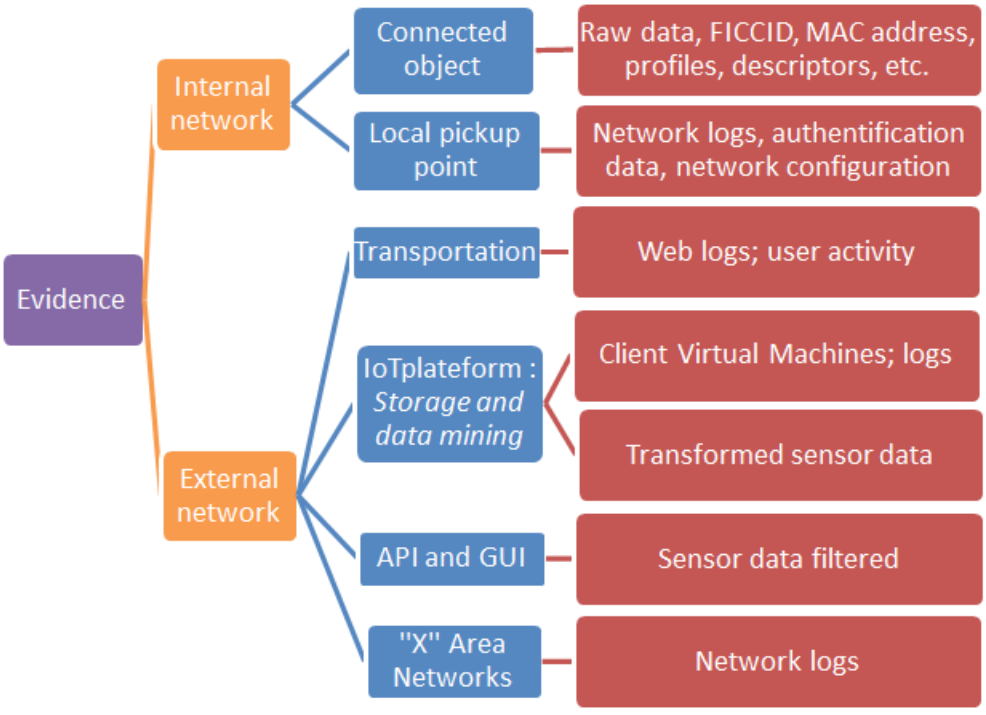


FIGURE 3.1 – Potentielles sources de preuves – Source : [4]

Les informations indirectes sont riches en enseignement. Elles orientent les investigations dans l'environnement criminel. Lorsqu'un incident se produit localement dans un dispositif

connecté, tous les journaux du flux de trafic constituent des preuves potentielles telles que l'on peut trouver dans les pare-feux ou les systèmes de détection d'intrusions *Intrusion Detection System* (IDS) [91]. Cependant, cette donnée doit-être contextualisée pour être valorisée. À titre d'illustration, une serrure intelligente est en mesure d'enregistrer les événements d'ouverture et de fermeture d'une porte. Par association, elle offre l'opportunité d'identifier la manière d'accéder dans un bâtiment et donc révèle la présence ou l'absence d'une personne. Elle date son passage. En reconnaissant le moyen de l'action (carte, téléphone portable, digicode, etc.), nous sommes en mesure de personnaliser l'événement et son potentiel acteur.

Face à la multiplicité des formats, comment étudier la donnée ? Faut-il effectuer un effort de priorisation ? Est-ce que toutes les données ont le même poids dans l'enquête ? Quelle stratégie d'analyse doit-être mise en œuvre par l'enquêteur pour valoriser au mieux cette information ? Existe-t-il des critères d'analyse et des liens cachés entre les données ?

3.2 Recherche des traces numériques dans l'Internet des objets

Dans le contexte de l'investigation « traditionnelle », les enquêteurs sont confrontés à des supports numériques classiques tels que des téléphones portables ou des ordinateurs. Ils sont formés pour rapidement les identifier et les collecter. Les supports méconnus sont, dans la plupart du temps, délaissés. Étant donné que les objets provenant d'une architecture de l'Internet des objets se présentent sous de nombreux formats, le processus d'identification devient plus complexe.

3.2.1 Identifier les sources potentielles dans l'environnement criminel

Les objets connectés de l'Internet des objets sont hétérogènes dans leur nature, leur usage et leur fonctionnement. Les plus connus d'entre eux sont les objets grands publics comme les équipements de santé ou de bien-être, les appareils ménagers, les assistants vocaux, la domotique, etc. Cependant, de nombreux objets connectés proviennent également de l'agriculture, de l'industrie, des transports ou des grandes infrastructures. De plus, les phénomènes du « fait maison » (*maker kit*) et du « DIY » (*Do It Yourself*) se développent. Ils passent notamment par la transformation d'éléments divers en objets connectés par l'apposition de moyens de communication, comme avec le cas d'un végétal dit connecté. Ainsi, les objets rencontrés par les enquêteurs ne sont pas toujours reconnaissables immédiatement, en raison d'une absence d'éléments d'identification normalisés ou de moyens permettant de déterminer une quelconque

connectivité. Certains dispositifs demeurent cachés sur la scène de crime et émettent potentiellement une faible signature. Nous pouvons citer l'exemple d'une caméra embarquée dans un ours en peluche ou celui d'un déodorant contenant un système de stockage Wi-Fi avec des données à caractères pornographiques mettant en scène des mineurs. Ces situations proviennent de perquisitions domiciliaires. La recherche et l'identification des objets connectés apparaissent comme un enjeu important pour le succès de l'enquête judiciaire. Elle doit donc faire appel à une succession d'opérations dans la détection, la localisation, la reconnaissance des objets et le recoupement de l'information.

L'objet est souvent sélectionné en fonction de ses aspects mécaniques et de la proximité directe par rapport à l'événement survenu. Néanmoins, le système connecté constitue dans de nombreuses situations un acteur passif, mais tout aussi pertinent. C'est le cas pour les thermostats connectés disposant de la fonctionnalité de géorepérage (*geofencing*). Ces solutions sont développées pour configurer le déclenchement d'une action, telle que la régulation du chauffage d'une habitation. Cette initiative est opérée lorsque le téléphone portable de l'utilisateur est détecté par le thermostat. Du point de vue de la criminalistique, cette propriété technique traduit la présence du téléphone de l'utilisateur du service dans un champ proche du détecteur. Cette source d'information est exploitable lors de la contextualisation du fait criminel.

Comment identifier ou reconnaître efficacement un objet physique disposant d'une fonctionnalité de connectivité ? Existe-t-il des moyens pour le détecter et le localiser dans l'environnement ? Comment différencier chaque équipement physique ? Comment établir le lien de dépendance qui réunit deux dispositifs pour un service connecté ?

3.2.2 Appréhender l'environnement connecté

L'opération de collecte consiste à extraire une donnée et/ou son contenant d'un environnement local vers un espace sécurisé garantissant leur intégrité. La diversité des objets, leurs rôles et leur dépendance à l'égard du réseau rendent cette phase difficile et périlleuse. L'Internet des objets se développe autour de plusieurs typologies de réseau, avec plus ou moins de dépendances. Chaque écosystème dispose d'une politique de gestion de la donnée propre, que ce soit au niveau du stockage, mais également que ce soit dans la remontée de l'information au sein du réseau. À ces difficultés, s'ajoute la problématique de la dispersion de la donnée au sein de l'infrastructure, dans l'environnement local et en ligne.

Les objets saisis ont la particularité de communiquer avec l'extérieur. Certains matériels ne peuvent pas être éteints sans perdre de l'information. Comment appréhender avec efficacité l'environnement et sa connectivité ? Est-ce que l'enquêteur doit nécessairement procéder à une extraction de l'information utile sur l'environnement ou a-t-il la possibilité d'extraire les

contenants pour une analyse ultérieure ? Comment accéder aux données, compte tenu des nombreuses dépendances, de la dispersion et de la fragmentation des informations ? Comment collecter efficacement sans altérer les données et les objets ? Quelle crédibilité l'enquêteur doit-il accorder aux traces recueillies ? Sont-elles exploitables, fiables et solides pour l'enquête et donc présentables devant un tribunal ?

3.2.3 Analyser l'environnement connecté et ses données

La phase d'analyse est d'autant plus complexe que la donnée est dispersée et/ou fragmentée au sein de l'infrastructure connectée, localement, mais également en ligne. En l'absence de croisement des données, l'enquêteur se retrouve confronté à une information incomplète, imprécise ou n'appartenant pas au dispositif analysé. En effet, la donnée est susceptible d'être partielle dans l'objet connecté mais devient un ensemble cohérent dans l'arborescence numérique. Cette détermination de la présence et du positionnement de l'information demeure unique à chaque environnement. À cette problématique, s'ajoutent les dépendances au sein des écosystèmes par des « liens cachés ». Un même résultat admet plusieurs causes. L'allumage d'un équipement peut être engendré par un interrupteur physique, mais également par une application de gestion, par une programmation, par une action extérieure telle qu'une commande vocale ou une capture de phénomène. Comment aborder la phase d'analyse face à cette dispersion de la donnée ? Comment associer la bonne donnée au bon équipement ? Quels sont les critères de reconstruction des événements ? Comment établir les dépendances entre les matériels ? Quels sont les acteurs ayant été la cause d'un événement ? Quel est le chemin de la donnée dans l'infrastructure connectée ? Quel sens donner au résultat de l'analyse de la trace ?

3.3 Question de l'individualité de la trace

L'identification technique et la caractérisation des sources de preuve sont des défis majeurs à la criminalistique de l'Internet des objets. L'ensemble de l'enquête dépend de la nature de l'appareil connecté présent localement et de son ancrage dans l'infrastructure. Dans une démarche prospective, à la suite des travaux initiés sur l'appréhension fine de l'environnement connecté et son analyse, il est pertinent de réfléchir à la question de l'individualisation des traces et des équipements. Elle s'inscrit dans un besoin de classification automatique des objets numériques. L'empreinte digitale est l'une des caractéristiques les plus utilisées pour identifier et individualiser une personne (cf. chapitre 2 - principe de Kirk). Elle ne change pas au fil du temps et ses nombreuses variantes permettent son tri. Est-ce que la notion d'empreinte est déclinable dans un cadre numérique ? Est-il possible d'individualiser un dispositif connecté au

travers de ces caractéristiques techniques ? Avec quel niveau de granularité ? Est-ce que ces éléments sont exploitables dans l'automatisation du processus d'appréhension et d'analyse du dispositif ?

3.4 Exercice d'investigation sur une scène de crime

Pour illustrer nos travaux de thèse et répondre aux questions soulevées dans ce chapitre, nous proposons d'étudier une scène de crime contenant des dispositifs connectés. Le scénario de cet exercice est inspiré de faits réels, rencontrés au cours d'enquêtes judiciaires. Il concerne la découverte d'un cadavre dans un appartement.

3.4.1 Présentation du scénario

Le 10 avril 2018 à 8 heures, la police est alertée d'un cambriolage et de bruits d'arme à feu provenant d'un appartement. Une patrouille arrive à 8 heures 15 minutes sur le lieu des faits. Elle découvre la porte d'entrée de l'appartement forcée. L'endroit présente également de nombreuses traces de lutte et de violence. Des objets sont dispersés et brisés sur le sol. Lors de la reconnaissance des lieux, un corps sans vie d'une personne est trouvé, allongé sur un lit. Les enquêteurs mettent donc en œuvre les premières mesures de protection en gelant la scène de crime. Une équipe médico-légale, dont un spécialiste en informatique, prend en charge la scène de crime à 9 heures.

3.4.2 Présentation de l'environnement

L'appartement (cf. Figure 3.2) est d'une surface de 45 m^2. Il comprend trois pièces distinctes : une entrée (pièce 1), une chambre (pièce 2) et un salon (pièce 3). Il contient de nombreux objets connectés. Il est équipé d'un système domotique issu d'un kit *Orvibo*. Ce produit comporte deux capteurs d'ouverture (1 et 2) et un capteur de mouvement (3) couplé à une caméra Wi-Fi (4). Cette solution contrôle les deux ouvertures extérieures. Elle communique par le protocole ZigBee au travers d'une passerelle dédiée (5). Le système domotique est également constitué d'ampoules *Philips* (6 et 7) connectées à une passerelle (8). Elles sont situées dans les pièces 2 et 3 de l'appartement. Par ailleurs, quatre capteurs *cookie* de la marque *Sen.se* sont cachés dans des différentes pièces. Ils transforment les objets ménagers en objets connectés. Dans cette situation, les capteurs surveillent la température ambiante de l'appartement (9), le niveau d'eau de la machine à café (10), la position du vélo en extérieur (11) et la gestion de la bibliothèque (12). Tous ces objets sont reliés à une base *Sen.se Mother* par un protocole propriétaire (13). Ainsi, ces différentes passerelles *Orvibo*, *Philips* et *Sen.se*,

l'*Amazon Echo* (14), la *RaspberryPi0* (15) et la caméra *IP M136W* (16) sont connectées à Internet au travers d'un *WinkHub 2* (17).

La victime est allongée sur le lit de la chambre. Elle dispose d'une *Apple Watch 3* (18) au poignet du bras droit et d'un *iPhone SE* (19) dans sa poche. Caché dans le lit se trouve un capteur de sommeil *Terraillon Dot* (20). Parmi les autres objets présents dans l'appartement, il y a un *Sens'it* (21), un bracelet *Heroz* (22) et une balance *Nokia* (23).

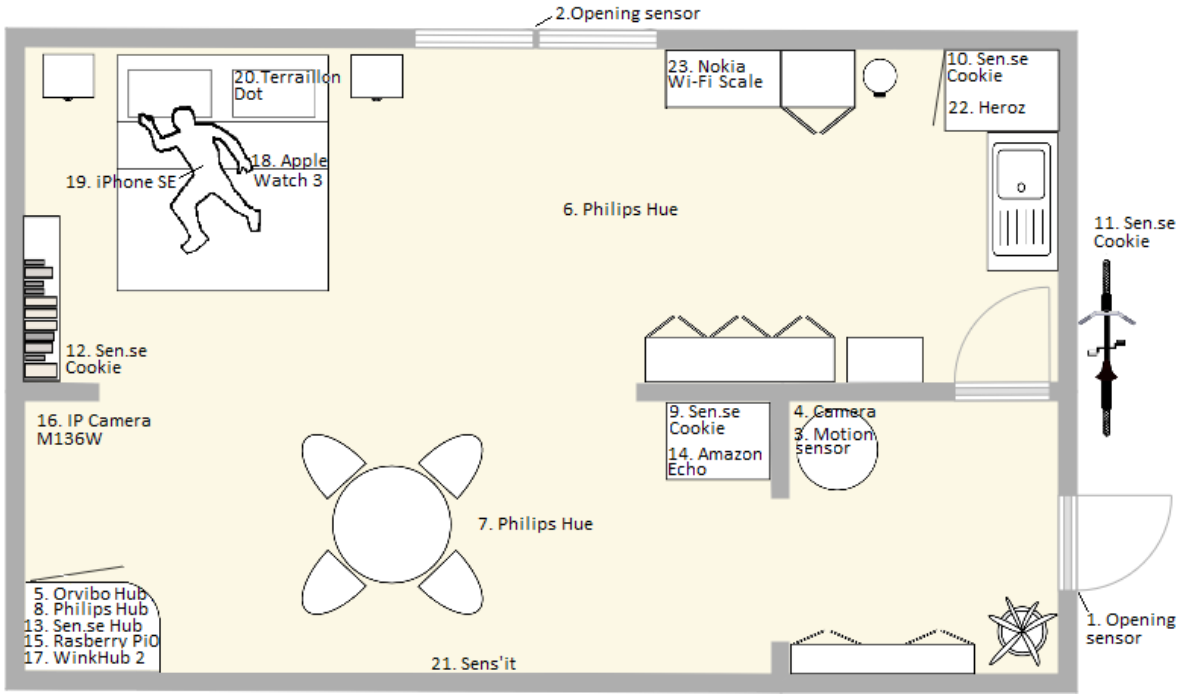


FIGURE 3.2 – Plan de l'appartement connecté de l'exercice d'investigation

3.5 En quelques mots : défi d'investigation dans l'Internet des objets

L'identification des objets, leur appréhension et leur caractérisation technique sont la clef de voûte dans le succès d'une enquête judiciaire dans l'Internet des objets. Or, la diversité d'usage, de fonctionnement relatif à la remontée et à la synchronisation de l'information ainsi que l'hétérogénéité des objets interdépendants rendent jusqu'à présent ce travail d'investigation très chronophage et fastidieux en l'absence de méthodologies et d'outils adaptés. Parcellaire dans un objet, la donnée prend son sens dans l'architecture globale.

Dans les prochains chapitres, nous apporterons des réponses aux problématiques successives dans la recherche des traces : l'appréhension avec intelligence des objets et de leur

environnement dans un contexte judiciaire (chapitre 4), la cartographie de la scène criminelle et de ses objets (chapitre 5), l'accès et la collecte de la donnée eu égard aux nombreuses dépendances (chapitre 6), leur exploitation par rapport aux causes de l'événement ainsi que la crédibilité données aux traces extraites (chapitre 7). Á ces questions se pose le besoin d'identifier et d'individualiser finement l'écosystème numérique et l'équipement connecté (chapitre 8).

Deuxième partie

Contributions

Chapitre 4

Classification des sources de traces numériques dans un écosystème connecté

L'Internet des objets est un réceptacle de traces de la vie quotidienne et de ses événements ponctuels. Son identification, sa caractérisation et sa classification conditionnent la réussite de l'exploitation des éléments de preuve dans le cadre d'une enquête judiciaire. Elles résultent de critères objectifs et techniques facilitant une lecture et une appréhension fidèle de l'environnement connecté.

Le NIST [92] définit la criminalistique numérique comme une « *science appliquée de l'identification, de la collecte, de l'examen et de l'analyse des données tout en préservant l'intégrité de l'information et en maintenant une chaîne de contrôle stricte pour les données* ». La phase d'identification se décline selon deux axes : une analyse de l'incident ou du phénomène survenu et une étude des sources de preuve.

4.1 Classification des objets physiques et de l'écosystème connecté

Il existe un large éventail d'applications de l'Internet des objets au regard des objets hétéroclites qui le composent. L'appréhension et l'analyse de cet écosystème pluriel passe par une classification basée soit sur l'objet primaire au travers de son domaine d'usage, ou bien soit à partir de la donnée émise et échangée constituant un élément de trace d'un point de vue de la criminalistique.

4.1.1 Objet au regard de son usage

De nombreux travaux proposent une catégorisation des objets à partir du domaine d'usage qu'il soit de l'ordre du personnel, de la maison, de l'entreprise, du service public ou de la mobilité [93]. Cette démarche s'adapte facilement aux familles d'objets disposant d'un périmètre d'emploi bien défini et unique. Cependant, certains dispositifs embarquent plusieurs fonctionnalités liées à plusieurs usages mettant à mal la définition stricte tels que les systèmes multimédia automobile ou les assistants vocaux. À cette difficulté s'ajoute la mutualisation des architectures connectées passant par des plates-formes et des services communs à plusieurs domaines d'application. Par ailleurs, les objets sont à même d'être détournés de leur fonctionnement pour lequel ils ont été configurés, créant de nouvelles valeurs d'usage. De façon plus générale, l'Internet des objets peut être abordé selon trois strates d'analyse, composées de l'électronique embarquée, de système et de système de systèmes. L'électronique embarquée est un système électronique et informatique autonome réalisant une tâche prédéfinie, parfois en temps réel. Le système se réfère à l'architecture autour des électroniques embarquées. Il regroupe l'ensemble des éléments interagissant entre eux, selon des règles prédéfinies. Le système de systèmes est un système où les éléments et les sous-ensembles sont eux-mêmes des systèmes. Il constitue donc un ensemble de systèmes autonomes interconnectés et coordonnés pour satisfaire une capacité ou des fonctions spécifiques que les systèmes indépendamment ne pourraient pas réaliser.

4.1.2 Objet au regard de la donnée

Dans le contexte d'une investigation judiciaire, il est intéressant de repartir des fondamentaux en explorant l'élément de preuve recherché afin de définir une classification des dispositifs physiques. Cette conception basée sur le type de données collectées est développée par Rahman et al [74]. Les auteurs proposent l'étude de scénarios avec l'usage de périphériques connectés dans la commission de l'infraction ; ici un *Sen.se Mother* et un *Samsung Hub*. L'idée est de réfléchir aux questions que les enquêteurs peuvent se poser au cours de leurs investigations.

Localement, les objets connectés collectent et génèrent des données d'environnement par le biais de différents capteurs ou actionneurs. Elles sont des captures instantanées d'un état physique de la scène infractionnelle. Les données sont catégorisables selon trois axes de lecture : un état contextuel, un positionnement dans l'espace et dans le temps d'une information, d'un objet ou d'un être vivant (cf. Figure 4.1). Le cas de la détermination d'une présence d'une personne sur une scène d'infraction illustre cette classification basée sur la donnée. Selon le principe de Locard, un individu laisse des traces lors de son passage sur un lieu. Les systèmes de détection sont en mesure de capter cet instantané. Cette information est également recoupée par des mesures d'environnement telles que des variations de température. Les éléments

recueillis offrent la possibilité de dater l'événement « présence » et de déterminer sa durée. Dans l'exercice de la scène de crime, les *Sen.se Cookies* sont utilisés à des fins multiples, comme le suivi du mouvement de l'objet vélo, avec l'heure et la durée du mouvement ou bien la mesure de la température ambiante, de l'usage de la bibliothèque ou de la machine à café. La signature des actions est capturée, analysée et reconnue par l'infrastructure *Sen.se* afin de délivrer une action spécifique, comme par exemple le déclenchement d'une alerte. Ces mesures sont la conséquence d'un changement d'état. Elles trahissent une présence ou un mouvement réalisé. La liaison radioélectrique est également considérée comme un capteur d'événement. A titre d'illustration, un changement de position d'un objet entraîne une restructuration de l'architecture du réseau basé sur le protocole *Routing Protocol for Low-Power and Lossy Networks* (RPL).

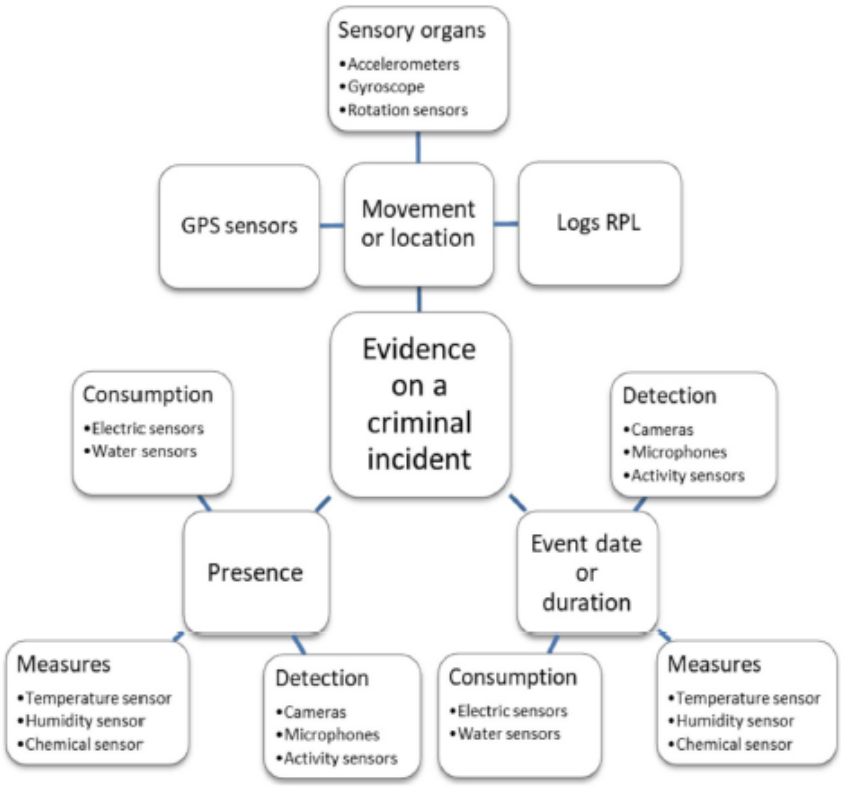


FIGURE 4.1 – Classification des objets connectés basée sur la donnée

La donnée locale est transmise à la plate-forme en ligne pour opérer une centralisation et une valorisation des données en service. Tout au long de ce processus, les différents acteurs de la chaîne numérique accompagnent et enrichissent la remontée de cette donnée brute avec une agrégation d'informations système, liées au fonctionnement et aux communications. Ces nouveaux éléments connexes à l'infrastructure complètent et améliorent la classification des équipements locaux.

4.2 Sélection des éléments de preuve dans une infrastructure IdO

La source de preuve est multiple. Il est donc primordial de définir des critères factuels de différenciation, mesurant la pertinence des données au regard de leur provenance.

4.2.1 Propriété de la donnée

Quatre caractéristiques principales basées sur les données guident la sélection des traces numériques : la pertinence, l'accessibilité, la localisation et le type de données. Chaque critère est illustré avec les matériels *Sen.se* et *Orvibo*. Cette taxonomie a été établie en s'appuyant sur les retours d'expérience d'experts dans l'étude forensique des équipements numériques.

4.2.1.1 Pertinence de la donnée

La pertinence des données se décompose en trois axes de lecture : la relation à l'événement, au temps et à l'espace. Une donnée est associée à une date de validité, contrainte par le temps. Elle est assujettie à une altération, à une perte de pertinence ou à une disparition. La contiguïté avec l'événement et le lieu de l'infraction jouent un rôle déterminant dans le choix des données à collecter. Face à l'événement, les appareils sont actifs ou passifs. Plus la proximité spatiale, temporelle et relationnelle est grande, plus les données collectées sont pertinentes et précises vis-à-vis de l'événement. La donnée est contextualisée par rapport au phénomène mesuré. Les faux positifs sont discriminés en présence de l'incohérence de situation ou en l'absence d'une vérification formelle. Dans l'appartement exercice, une action d'ouverture et de fermeture d'une porte est enregistrée par des capteurs de contact *Orvibo*. Cette mesure est souvent couplée à une détection de mouvement et à un déclenchement d'une action préétablie. Cependant, un simple claquement d'une porte voisine est susceptible de générer l'événement de capture. En l'absence de données de mouvement ou d'une redondance de mesure, cette donnée doit être pondérée et abordée de façon critique par l'enquêteur. Ce phénomène est également observable dans le cas de l'écosystème *Sen.se*, lors d'une perte de connexion entre la passerelle *Mother* et le capteur *Cookie*.

4.2.1.2 Accessibilité de la donnée

L'accessibilité des données présentes dans les appareils est également un critère pertinent dans la recherche de traces. La difficulté d'accès aux données est propre aux caractéristiques matérielles et logicielles du dispositif étudié. Les données sont également susceptibles d'être protégées par du chiffrement ou des techniques de protection contre la rétro-ingénierie. D'autre

part, certains médias jouent un rôle de catalyseur d'informations. Ils sont donc prompts à contenir plus d'informations pertinentes. Pour les capteurs *Cookie* ou *Orvibo*, la récupération de données passe par un accès matériel. Elle nécessite l'usage de solutions d'ingénierie avancée pour la lecture et le décodage. Les passerelles *Mother* et *Orvibo* disposent d'un accès Ethernet, des ports et des services de communication ouverts et exploitables lors de l'extraction.

4.2.1.3 Localisation de la donnée

La question de la localisation des données se concentre sur leur positionnement dans l'infrastructure connectée et sur la capacité de stockage des différents équipements. Elles sont soit stockées localement ou bien soit déportées sur des systèmes dédiés en externe. La plupart des appareils ne disposent pas de mémoire ou de grande capacité de stockage. Les informations sont placées dans des mémoires volatiles ou persistantes. Le positionnement des données dans l'infrastructure et les spécificités techniques des matériels ont un impact direct sur les actes techniques à exécuter par l'enquêteur. Pour accéder aux données contenues sur les plates-formes en ligne, une réquisition auprès des opérateurs est nécessaire. Cependant, cette action requiert de connaître les périphériques locaux et leurs caractéristiques techniques afin de fournir une demande précise et circonstanciée. Dans l'appartement exercice, les capteurs *Cookie* stockent environ 10 jours de données. Cette option est activée en cas de perte de connexion avec la base centrale *Mother*. Dès que la passerelle est accessible et opérationnelle, les matériels retransmettent l'ensemble des données. Cette dernière relaie l'information reçue à une plate-forme en ligne dédiée. Cet espace de stockage traite, interprète et analyse en permanence les informations reçues pour fournir des services aux utilisateurs. Les données collectées sont accessibles par une interface Web ou une application installée sur un téléphone portable. Les enquêteurs sont en mesure d'extraire des données des capteurs *Cookie*, les journaux d'événements présents sur la passerelle *Mother* et les informations centralisées dans l'application mobile *Sen.se*. Á partir de ces informations locales, ils requièrent auprès de l'opérateur *Sen.se* les données présentes sur ses plates-formes.

4.2.1.4 Type de la donnée

Le type de données est défini selon trois grandes caractéristiques : directe, mécaniquement transformée et interprétée par l'homme. Les données directes sont des données brutes mesurées par un objet. Les informations sont contextualisées par les dispositifs en fonction des paramètres techniques tels qu'une notion de seuil. L'analyse des journaux d'événements et l'observation de l'évolution des données au fil du temps constituent les données indirectes. Elles reflètent l'état de fonctionnement des objets. Ainsi, une présence humaine est déduite d'un événement enregistré résultant de la manipulation d'un interrupteur dans une pièce.

L'événement est associé à un lieu fixe et nécessite une intervention physique pour être activé. Les données directes et transformées sont plus facilement exploitables que les données interprétées. Cette analyse se base sur une connaissance approfondie des dispositifs dans leur fonctionnement nominal. Par exemple, le capteur *Cookie* collecte des données brutes de son environnement telles que la température ou la détection de présence. Ces éléments sont facilement exploitables en l'état par les enquêteurs à une conversion près en ce qui concerne la température. Un changement de donnée renvoie à un événement factuel. Pour des raisons d'économie d'énergie, les informations transmises sont asynchrones. Leurs valeurs sont arrondies et moyennées en fonction des besoins de précision et de la criticité des mesures attendues. Les journaux de connexion trouvés sur la passerelle *Mother* retracent la chronologie des échanges réalisés.

4.2.2 Pondération des sources IdO

Á partir des propriétés décrites dans les sections précédentes et la pertinence de la collecte d'information, l'infrastructure globale est évaluée. Cette classification met en relation des critères objectifs sur la donnée en tenant compte des conditions de coût, de performances et de critères de qualité. L'enjeu est de déterminer les meilleures sources de preuves potentielles au profit de l'enquête judiciaire. Les quatre colonnes de la table 4.1 sont définies par les critères sur la donnée. La rangée se divise selon un découpage en quatre sections de l'IdO, soit l'objet connecté, la passerelle, la plate-forme en ligne et l'interface de l'utilisateur. Ces différents éléments sont déclinés au regard des notions suivantes :

- la productivité, en l'effort engagé pour obtenir les données attendues ;
- le coût humain, en un coût lié au temps d'exécution des opérations techniques ;
- les coûts d'ingénierie, en un coût financier engagé dans le succès de l'opération ;
- l'altération, en l'impact des opérations techniques sur les appareils et les données extraites.

Le critère 1 correspond au poids le plus fort et 4 au poids le plus faible.

Nous allons maintenant examiner plus en détail les pondérations à partir du cas d'usage *Sen.se Mother* provenant de notre scénario d'exercice décrit dans le chapitre 3.

		Pertinence	Accessibilité	Position	Type	Total	#
Productivité	Objet connecté	1	4	4	3	12	2
	Passerelle - Nœud	4	1	3	4	12	2
	Plate-forme	2	3	1	2	8	1
	HCI (API - GUI)	3	2	2	1	8	1
Coût humain	Objet connecté	4	4	3	4	15	3
	Passerelle - Nœud	3	1	1	3	8	1
	Plate-forme	1	2	4	1	8	1
	HCI (API - GUI)	2	3	2	2	9	2
Coût d'ingénierie	Objet connecté	4	4	3	1	12	3
	Passerelle - Nœud	3	2	1	4	10	2
	Plate-forme	1	1	4	3	9	1
	HCI (API - GUI)	2	3	2	2	9	1
Altération	Objet connecté	4	4	3	1	12	3
	Passerelle - Nœud	3	3	2	4	12	3
	Plate-forme	1	1	4	3	9	2
	HCI (API - GUI)	2	2	1	2	7	1

TABLE 4.1 – Classification de l'infrastructure de l'Internet des objets, basée sur les propriétés de la donnée au regard des contraintes techniques et opérationnelles

4.2.2.1 Question de productivité

La question de la productivité est illustrée par la mesure de la température ambiante par un capteur *Cookie*. Cet objet connecté repose sur un système de développement complet basé sur le microcontrôleur TI CC430, qui fournit un capteur thermométrique intégré. Son impédance équivalente est de 51 kΩ. La mesure est effectuée en Fahrenheit avec une précision d'acquisition d'un centième de degré. L'acquisition des données nécessite le dessoudage et la lecture de la puce électronique. Cette dernière ne contient que des mesures récentes de la température. La passerelle *Mother* contribue uniquement à la transmission des données de la température vers la plate-forme en ligne *Sen.se*. Elle contient cependant les journaux d'événements retraçant les connexions et les interactions avec le système. Elle n'est pas très sécurisée. Elle comporte de nombreux ports ouverts avec leurs services de communication. L'interface utilisateur *Sen.se* offre des informations centralisées, formatées et interprétées. Toutefois, les données ont été sélectionnées et mises en forme en fonction des choix du concepteur de l'application et de ses versions de développement. L'interface renvoie les données de température

mesurées sur une durée longue, sous la forme d'un graphique affiché avec une précision de 0,1 degré. Leur accessibilité dépend du conteneur : un téléphone portable ou un portail Web avec une authentification. En fonction des caractéristiques du téléphone, des sécurités avancées peuvent être activées nécessitant l'usage d'outils criminalistiques adaptés. La plate-forme en ligne contient la base de données du fabricant *Sen.se*. Cette solution centralise et enregistre les informations brutes remontées par les équipements locaux (*Cookie* et *Mother*). Ainsi, la précision des données remontées est de l'ordre d'un centième de degré. La collecte des données nécessite l'intervention de l'opérateur privé ou de prestataires externes. Elle est conditionnée par la qualité des éléments renseignés dans la réquisition.

Deux erreurs de conversion sont possibles dans l'écosystème *Sen.se*. La première peut se produire dans le *Convertisseur Analogique-Numérique* (CAN) 12 bits du microcontrôleur lors de la conversion de la mesure électrique en données de température. Le décalage du capteur de température peut être important. La structure du descripteur du dispositif contient des valeurs d'étalonnage de 30°C ±3°C et 85°C ±3°C pour chacun des niveaux de tension de référence. La seconde est susceptible de se produire lors de la conversion des données de température dans l'application, de degrés Fahrenheit en degrés Celsius. Cette imprécision doit être prise en compte lors de la phase d'analyse dans l'évaluation de la qualité de la trace collectée. Elle est d'autant plus vraie lorsque la mesure est critique par rapport au succès de l'enquête.

4.2.2.2 Question du coût humain et d'ingénierie

De nombreux équipements interconnectés sont présents sur la scène de crime. Chaque dispositif est unique et implique des coûts importants de recherche et de développement en matière d'extraction et d'analyse des informations. Cette contrainte est d'autant plus forte dès lors que les formats de données sont propriétaires. Le capteur *Cookie* contient un microcontrôleur CC430F6137 avec 32 KB de mémoire flash programmable dans le système et 4 KB de RAM. Ce composant dispose d'un accélérateur *Advanced Encryption Standard* (AES) de 128 bits pour sécuriser l'échange de données. En ce qui concerne les passerelles, le principal coût est lié à l'interprétation des informations collectées. L'écoute des ports détermine ceux qui ouverts et exploitables dans l'extraction de la mémoire. La *Mother* dispose d'une connexion Ethernet pour échanger de l'information avec l'infrastructure de l'IdO. Elle utilise les ports TCP 123, 443, 6514, 8482 et *User Datagram Protocol* (UDP) 53, informations recueillies avec l'outil d'identification et de pré-analyse présenté ci-après. Sa mémoire est également accessible et lisible par dessoudage, en l'absence de chiffrement. Elle est composée d'un noyau Linux dans lequel les journaux de connexion et la configuration de l'infrastructure sont conservés. Les coûts financiers et humains résultant des réquisitions [2] auprès des opérateurs de

2. Les réquisitions informatiques sont prévues par l'article 60-2 du *Code de procédure pénale* (CPP).

plate-forme sont partiellement contrôlés. Ils sont définis par un contrat entre les parties. Les demandes transmises aux opérateurs doivent cependant être précises afin d'obtenir une information pertinente et exploitable par les enquêteurs. La réquisition doit au minimum contenir des informations techniques relatives à l'identification de la passerelle sous la forme d'une adresse *Medium Access Control* (MAC) ou d'un numéro de série du modèle. Pour l'interface utilisateur, le principal coût résulte dans l'accès et l'extraction des données de l'application. Toutefois, cette approche repose sur des processus maîtrisés dans le cadre de la criminalistique numérique en téléphonie.

4.2.2.3 Question de l'altération

Pendant les phases de collecte, les objets sont susceptibles d'être endommagés. Les données sont dans ce cas corrompues ou perdues définitivement. Les méthodes matérielles, telles que le dessoudage, le *Boundary Scan* ou le *Test Access Port* (TAP), sont très destructrices mais offrent un plein accès aux données présentes dans les mémoires de stockage. L'information extraite est potentiellement chiffrée. L'approche logicielle, moins préjudiciable pour l'objet, modifie son fonctionnement nominal. Elle donne accès à une donnée déchiffrée mais nécessite un accès privilégié au système. Ainsi, elle est susceptible de générer une écriture sur celui-ci. Les démarches d'extraction des données sont développées dans le chapitre 7 de la thèse. Dans le cadre de l'exercice, un capteur *Cookie* dispose de 4 Ko de RAM. Ce stockage est utilisé comme mémoire tampon lorsque la transmission de données est interrompue. Une mise hors tension de l'équipement entraîne une perte d'information. L'accès à sa donnée nécessite, en l'occurrence, l'usage d'une approche matérielle. En l'absence de moyen nécessaire à la réussite de l'extraction, l'objet est exploité ultérieurement. Il doit donc être maintenu dans un état garantissant sa non-altération tout au long du processus de collecte et de placement sous scellé. Ces processus font l'objet d'un approfondissement dans le chapitre 6. L'approche logicielle est également mise en œuvre pour l'extraction des données en ligne, dans le cadre d'une perquisition « *en ligne* ». Elle doit cependant respecter le cadre légal en particulier dans l'exploitation d'un client d'accès [3]. L'extraction des données contenues dans l'application et sur les passerelles locales combine les deux approches en fonction du contenant à exploiter et de l'information recherchée pour les besoins de l'enquête.

3. La perquisition « *en ligne* » est définie par la loi du 18 mars 2003 et dépend du cadre légal. Elle est prévue par l'article 57-1 du CPP pour l'enquête de flagrance et l'article 76-3 du CPP pour l'enquête préliminaire.

4.2.2.4 Limites de la classification proposée

À partir des différentes pondérations, une classification des dispositifs de l'infrastructure IdO est définie (cf. Tableau 4.2).

	Total par donnée	#
Objet connecté	51 *(12+15+12+12)*	4
Passerelle - Nœud	42 *(12+8+10+12)*	3
Plate-forme	34 *(8+8+9+9)*	2
HCI (API - GUI)	33 *(8+9+9+7)*	1

TABLE 4.2 – Bilan de la classification

Selon l'ordonnancement obtenu, l'interface homme-machine (IHM) constitue la partie la plus performante en termes de source potentielle de preuves. En raison de sa position stratégique, l'interface demeure un point de convergence des données provenant des dispositifs connectés. Les éléments reçus sont interprétés. Ainsi, ils sont facilement exploitables par les enquêteurs pendant l'analyse. Néanmoins, l'application ne contient que les données que son concepteur logiciel a souhaité retourner. Par exemple, les informations liées au fonctionnement de l'infrastructure connectée n'apparaissent pas nécessairement. La plate-forme en ligne offre également des performances intéressantes grâce à son rôle de concentrateur de données. Elle contient des informations supplémentaires par rapport à l'interface applicative, en particulier, des éléments de fonctionnement à des fins de maintenance et de gestion des incidents. Cependant, les données sont déportées à l'extérieur de la scène de crime. Leur accès dépend de l'intervention d'une tierce personne à l'enquête et de l'établissement d'accords de collaboration entre l'exploitant et les autorités judiciaires. Ainsi, la manière de collecter les informations est indépendante de la volonté de l'enquêteur. La passerelle contient principalement des informations indirectes sous la forme de journaux. L'exploitation de ces éléments nécessite une bonne connaissance de l'architecture et du fonctionnement du réseau. Les objets connectés constituent l'interface directe avec la scène de crime. Ils sont les témoins et les acteurs des événements survenus. Ils contiennent un instantané brut du phénomène capté. Ainsi, ils ne contiennent que des informations localisées et spécifiques. Ils n'ont pas de vision globale sur les données échangées dans l'infrastructure. Leur exploitation est pertinente lorsque l'enquêteur recherche une donnée précise et locale qui n'a pas été synchronisée avec le réseau. C'est notamment le cas pour une montre d'un joggeur ou un analyseur de sommeil déconnecté de son infrastructure. Ces équipements connectés contiennent une information non remontée en cas de dissociation avec leurs passerelles respectives.

La priorisation des sources d'information est remise en question en fonction du contexte de l'incident, comme dans le cas d'une absence de synchronisation entre les objets et le réseau local. L'architecture est parfois conçue selon un principe de l'information géodistribuée (*fog computing*) ou en périphérie (*edge computing*). Les ressources de calcul et d'analyse sont réparties localement entre la source et le Cloud. Seules des données traitées sont transmises en ligne. Dans ce cas, les éléments bruts sont hébergés par les nœuds ou par les objets connectés. Certaines formes d'environnement n'intègrent ainsi pas de partie *Cloud* et donc de solutions de traitement déportées. Elles sont structurées en boucle fermée autour de liaisons locales regroupant des objets connectés, des passerelles et des interfaces de commande. Cette structuration répond souvent à des problématiques de sécurité. De ce fait, cette classification reste générique et doit s'adapter aux caractéristiques des dispositifs connectés, des conditions d'environnement et des besoins de l'enquête. L'identification des équipements connectés est donc primordiale pour définir leurs propriétés de fonctionnement dans l'écosystème connecté.

4.3 Éléments caractéristiques d'un équipement communicant

Le processus de reconnaissance des équipements est effectué à partir des informations visibles sur le produit et/ou son électronique. Des tables de correspondance permettent d'effectuer les rapprochements nécessaires. Elles sont enrichies par les recherches réalisées sur les objets connectés, par les connaissances scientifiques dans le domaine de l'Internet des objets et par les retours d'expérience des enquêtes.

4.3.1 Étude des caractéristiques visuelles

Les équipements émetteurs d'ondes sont plus ou moins renseignés selon la réglementation du pays d'usage ou de fabrication. Ils peuvent porter les mentions suivantes : une marque du constructeur ou de produit, un numéro de modèle ou de série, un identifiant unique, un lieu de fabrication, etc. (cf. Figure 4.2). Ces informations sont également susceptibles d'évoluer d'un pays à un autre.

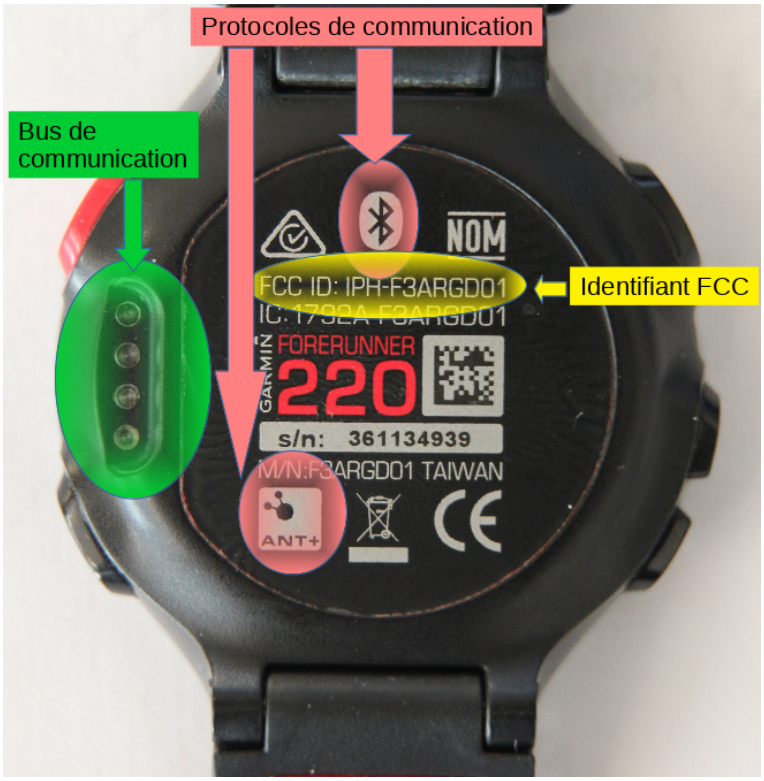


FIGURE 4.2 – Exemple de caractéristiques visuelles

L'étude des équipements présents dans l'exercice de la scène de crime montre que tous les objets ne disposent pas des mêmes inscriptions (cf. Tableau 4.3). Bien que ces inscriptions diffèrent, certains éléments sont redondants d'un matériel à l'autre. Une famille d'objets est identifiable par un nom de fabriquant et de modèle, un numéro de modèle et un identifiant unique. Cette dernière information renvoie à des bases de données contenant les spécifications et des caractéristiques des objets. Les plus utilisés sont le *Federal Communications Commission* (FCC), le *China Minitry of Industry and Information Technology* (CMIIT), le *Korean Communications Commission/ Ministry of Science, ICT and Future Planning* (KCC/MSIP), le *Industry Canada* (IC) et le *Agency of National TELecommunications* (ANATEL). Ces bases sont plus ou moins renseignées et normalisées, en fonction des législations des pays et des fabricants. Un objet unique est quant à lui identifiable par une adresse physique, comme par exemple une adresse MAC ou un numéro de série *User IDentifier* (UID). Ces éléments sont souvent inscrits sur la carte électronique ou dans le logiciel. Son accès nécessite un démontage de l'équipement comme dans le cas de la passerelle *Philips* ou une étude poussée des communications pour les capteurs *Cookie* où chaque capteur est identifiable individuellement par un nom composé de huit lettres et chiffres, tels que 6FEB205D ou AB61092C. L'étude de

cette signature individuelle permet à l'enquêteur de connaître le nombre d'objets connectés présents sur le lieu du crime.

No.	Type d'équipement	Nom du constructeur	Numéro de modèle	Identifiants (FCCID, etc.)	Lieu de fabrication	Nom de produit	Protocole utilisé	Numéro de série	MAC
1 - 3	OC	X	X	X	X	X	-	-	-
4	OC	X	X	X	X	-	-	-	-
5	P	X	X	X	-	X	X	-	-
6 - 7	OC	X	X	X	X	X	-	X	-
8	P	X	X	X	X	-	X	-	X
9 - 12	OC	X	X	X	X	-	-	-	-
13	P	X	-	X	X	X	-	-	-
14	OC et P	X	X	X	X	-	-	-	-
15	OC et P	X	X	X	-	X	-	X	-
16	OC	-	-	-	-	-	-	-	-
17	P	X	X	X	X	X	X	X	X
18	OC	X	X	X	X	X	X	-	-
19	P et API	X	X	X	X	X	-	-	-
20	OC	X	X	X	X	X	-	X	-
21	OC	X	X	X	X	X	X	X	-
22	OC	-	-	-	X	-	-	-	-
23	OC	X	X	X	X	X	-	-	-
Total (OC, P et API)		89%	83%	89%	83%	67%	28%	28%	11%
	Selon OC (13)	85%	85%	85%	85%	46%	15%	31%	0%
	Selon P (7)	100%	86%	100%	71%	71%	43%	29%	29%

Légende : 1-3 : Capteur Orvibo — 4 : Caméra Orvibo — 5 : Passerelle Orvibo — 6-7 : Ampoule Philips — 8 : Passerelle Philips — 9-12 : Sen.se Cookie — 13 : Sen.se Mother — 14 : Amazon Echo — 15 : RaspberryPi0 — 16 : IP Camera M136W — 17 : WinkHub 2 — 18 : Apple Watch 3 — 19 : iPhone SE — 20 : Terraillon Dot — 21 : Sens'it — 22 : Bracelet Heroz — 23 : Balance Nokia — OC : Objet Connecté — P : Passerelle

TABLE 4.3 – Étude visuelle des appareils présents sur la scène de crime

4.3.2 Outil d'identification à partir des éléments visuels et de la pré-analyse protocolaire

Afin d'automatiser l'identification des équipements, une interface est développée en Python 3 avec une base SQLite (cf. Figure 4.3 et 4.4). Elle fait correspondre les informations observées avec des données recensées dans les bases d'identification. Elle s'appuie sur les identifiants des familles de produits et l'adresse physique MAC. Les autres critères n'ont pas été retenus en l'absence de moteurs en ligne offrant des résultats suffisamment filtrés ou pertinents.

En raison des contraintes opérationnelles des unités, l'outil s'appuie sur une base locale incrémentée par des informations provenant des bases en ligne (cf. Figure 4.5). La récupération des données se fait par *web scraping*.

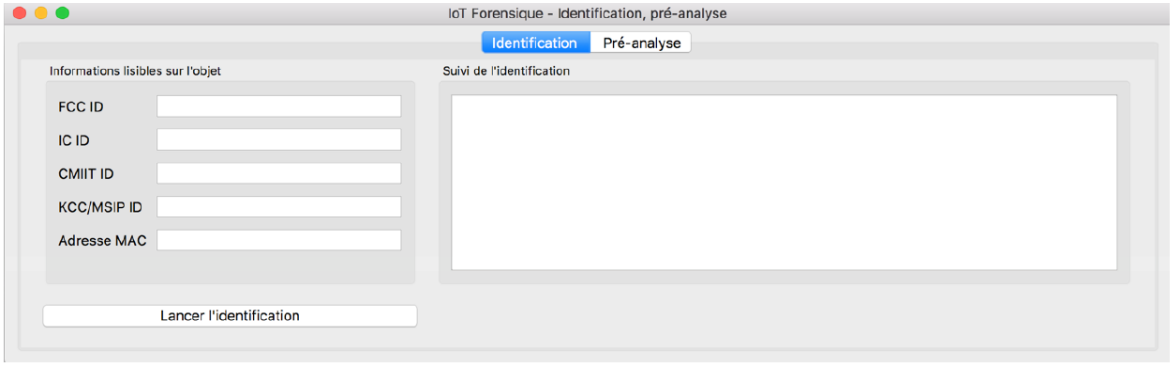


FIGURE 4.3 – Conception de l'outil d'identification

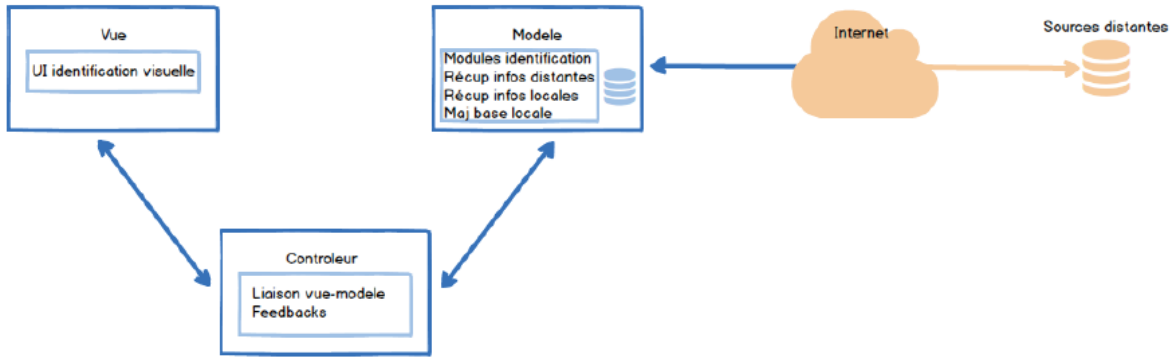


FIGURE 4.4 – Vue MVC de l'outil d'identification

Dans le cas d'usage, l'identifiant FCC de la passerelle *Philips* est 2AGBW3241312018AX, son IC 20812-2018X, son CMIIT 2016DP3836 et son KCC/MSIP MSIP-CRM-pli-3241312018A. Le résultat retourné par l'outil d'identification est présenté en figure 4.6. Il renseigne l'enquêteur sur le type d'équipement ainsi que les protocoles de communication et les plages de fréquence associées à la passerelle.

L'outil embarque également une fonctionnalité de pré-analyse (cf. Figure 4.7) basée sur *Nmap* pour les équipements disposant d'une connectique RJ45, impliquant l'utilisation du protocole Ethernet, ainsi que des autres protocoles réseau « classiques » (TCP, IP, *Address Resolution Protocol* (ARP), etc.). Ce développement s'appuie sur le principe que les implémentations de ces protocoles prévoient des moyens permettant d'identifier les machines d'un réseau (adresse IP, adresse MAC, fonctions de discovery, envoi automatique de paquets beacon pour signaler leur présence, etc.) afin d'assurer les fonctionnalités des équipements (maintenance des tables de routage, des tables ARP, etc.). Il est possible d'utiliser ces techniques pour identifier le matériel. Les solutions passives sont privilégiées contenu de manière à limiter l'altération des dispositifs étudiés.

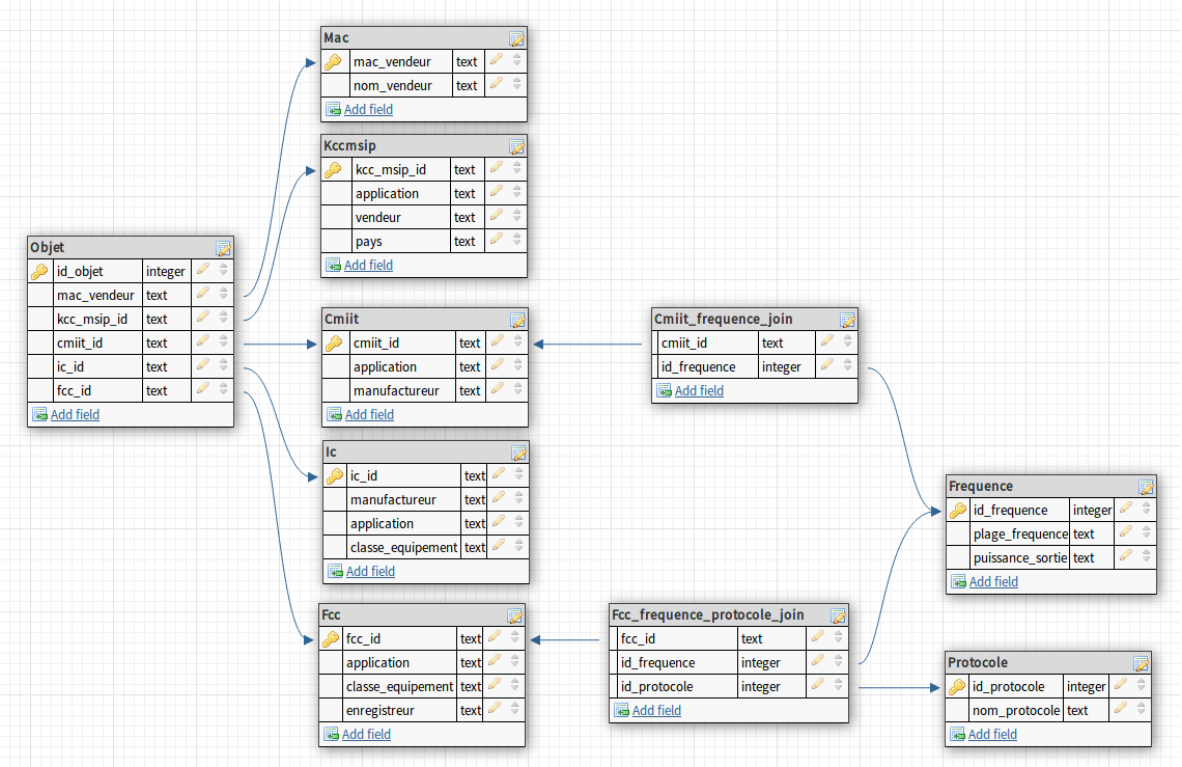


FIGURE 4.5 – Structure de la base de données

Cette pré-analyse donne également des informations aux enquêteurs pour définir les méthodes d'acquisition des dispositifs par l'exploitation de services actifs fournissant un accès distant (protocole *Secure Shell* (SSH), *Terminal network ou Telecommunication network* (Telnet), *File Transfer Protocol* (FTP), serveur web, etc.).

4.3.3 Recoupement des informations recueillies

Une phase de recoupement de l'information finalise le processus d'identification. L'enquêteur compare les données techniques de la base de connaissance et de l'environnement observé sur la scène de crime. Il attribue pour chaque matériel retrouvé une fonction dans l'infrastructure connectée : objet connecté, nœud du réseau local, passerelle, IHM, ainsi que les différentes dépendances au réseau. Il procède à l'élaboration de sa cartographie sous la forme d'un graphique de dépendances. Cette opération permet de vérifier la cohérence des informations recueillies et de déterminer celles manquantes. Par exemple, la passerelle *Mother* communique avec les capteurs *Cookie* sur le réseau sans fil, à partir d'un protocole propriétaire en 915 *Mégahertz* (MHz), *dixit* les informations renseignées en la base. Cette passerelle utilise une connexion filaire permettant de communiquer vers l'extérieur. L'identification de la box

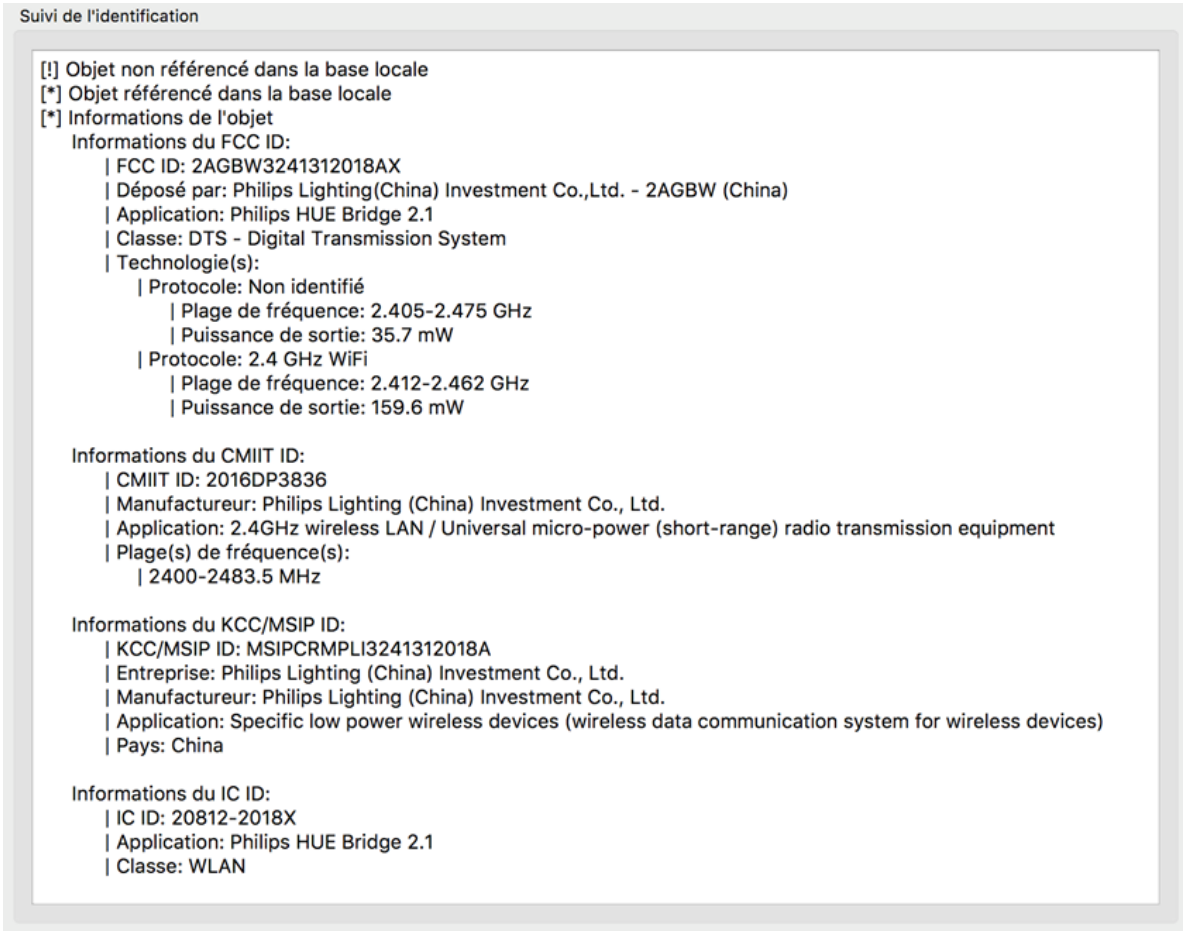


FIGURE 4.6 – Résultat d'une identification d'un objet non référencé localement

Mother et des paramètres du réseau donne des informations utiles sur la plate-forme en ligne *Sen.se*, telle que l'identité et les coordonnées de l'opérateur.

4.4 Classification et identification à l'épreuve d'un cas d'usage

Dans cette section, la classification et le processus d'identification visuelle est questionnée par rapport à l'objet connecté de santé, *Terraillon Dot*. Les autres objets présents sur la scène de crime sont ignorés afin de faciliter l'étude. Cet équipement est découvert à proximité du corps de la victime. Lorsque l'enquêteur judiciaire le trouve, les causes et le contexte du décès ne sont pas définis : homicide ou suicide. Aucun élément matériel ne permet de privilégier une piste ou une hypothèse de travail.

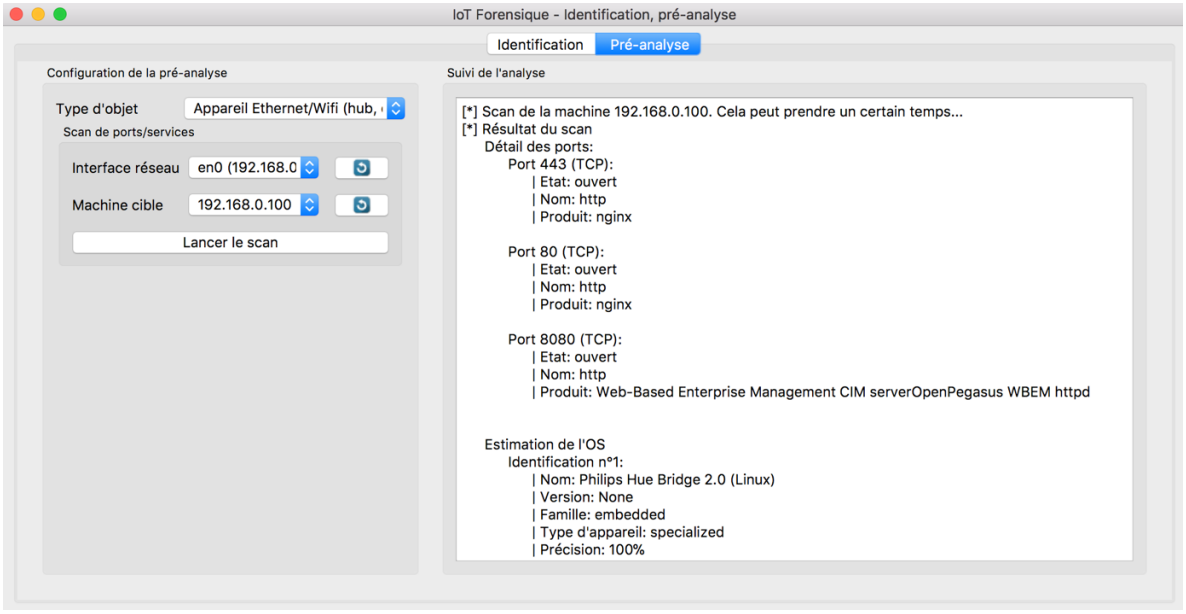


FIGURE 4.7 – Interface de l'outil de pré-analyse réseau

L'objet connecté est identifié par le symbole *Terraillon*, avec un FCC ID 2ADIOB501, CMIIT ID 2016DP2433, le nom du modèle B501 et un numéro de série 17302162. D'après les informations retournées par la base de données techniques, l'objet connecté est un capteur de sommeil *Sleepace Dot*, de 3,3 cm de diamètre et de 1,3 cm d'épaisseur. Cet équipement communique en Bluetooth 4.0 Low Energy, sur la gamme de fréquence 2,402-2,48 *Gigahertz* (GHz) avec une puissance de sortie de 1,1 *Milliwatt* (mW). Il émet dans un rayon maximal de 10 mètres (33 feet) en intérieur. Le *Sleepace Dot* utilise une passerelle pour communiquer avec Internet. L'étude de la scène de crime permet de trouver un *iPhone SE* potentiellement associé au dispositif. Cependant, aucun élément ne permet de corroborer cette opinion en l'état. Cet élément ne peut être confirmé ou infirmé en l'absence d'une analyse des journaux d'événements et des données de l'application de contrôle-commande.

L'analyse des deux équipements est opérée selon la classification énoncée précédemment. L'application *Wellness Coach* fournit des informations personnelles sur l'utilisateur, ses activités et le fonctionnement de l'objet connecté. Elle contient les horodatages du déclenchement du capteur et les mesures des mouvements de l'utilisateur au repos pendant plusieurs mois. Cependant, seules les informations synchronisées manuellement sont disponibles dans l'application. L'association entre l'*iPhone* et le *Terraillon* est définie à partir des journaux d'événements de la communication Bluetooth du téléphone et des données de synchronisation de l'application. Les dernières heures d'enregistrement précédant l'événement n'ont pas été remontées à l'application en l'absence d'une synchronisation manuelle par l'utilisateur. Les données de la plate-forme *Terraillon* sont similaires aux informations recueillies sur l'applica-

tion mobile. Le cloud ne contient donc pas de données sur les derniers instants de mesure. Le capteur de sommeil dans ce cas joue un rôle décisif dans la résolution de l'enquête criminelle. En l'absence de remontée de données dans le système d'information, l'objet connecté est capable de contenir les mesures d'une semaine d'activités. Á partir des données extraites de cet objet, nous reconstituons l'ensemble de la chronologie des événements, en datant les habitudes de vie de la victime, les derniers instants de mesure et les interactions avec l'environnement.

Dans ce cas précis, nous nous sommes focalisés sur l'étude d'un unique objet connecté. Cependant, les scènes de crime sont susceptibles de contenir de nombreux autres objets interconnectés offrant des informations cruciales pour l'enquête. Par exemple, le *Dot Sleep Sensor* peut être couplé à une solution *Homni Smart Standby* contenant des capteurs environnementaux (température, humidité, niveau sonore et luminosité) et à la solution *Reston* pour calculer le rythme cardiaque et respiratoire d'une personne. Il est donc toujours pertinent de partir de l'étude des données présentes dans l'application téléphonique *Wellness Coach - Sleep*, en raison de son rôle de catalyseur des informations. Selon les besoins d'enquête, certains matériels peuvent faire l'objet d'une étude plus approfondie. Cependant, la probabilité de trouver des informations pertinentes et complètes dans le capteur isolé est plus faible que dans l'IHM. Le chapitre 7 détaille les études spécifiques qui nous ont permises d'établir rigoureusement la validité de ces éléments.

4.5 En quelques mots : besoin d'identifier et de classer les sources de traces numériques dans un écosystème connecté

Étant donné le volume considérable de données à étudier, les enquêteurs doivent trouver des processus et des instruments efficaces pour les sélectionner et les caractériser. Ils ne peuvent pas tout collecter et tout analyser en raison de contraintes opérationnelles et économiques. L'identification visuelle est opérée dès le début de l'enquête afin d'optimiser le processus de la collecte. Elle s'appuie sur une panoplie d'outils dans la recherche et la caractérisation des équipements.

Chapitre 5

Recherche des sources de traces numériques dans un écosystème connecté

La recherche des éléments de preuve sur une scène de crime est une opération réalisée lors de la phase d'identification. Elle consiste à appréhender et à caractériser l'environnement physique afin, dans un second temps, de procéder à la collecte et à l'analyse des traces numériques. Dans le cadre de l'Internet des objets, elle revêt une importance particulière face à des dispositifs difficilement identifiables et polymorphes.

5.1 Internet des objets et localisation

L'appréhension des dispositifs connectés répond à des exigences techniques et juridiques liées à l'environnement étudié et aux informations collectées.

5.1.1 Défi de l'identification des dispositifs connectés

Lors d'investigations et de perquisitions judiciaires, les enquêteurs sont formés dans la recherche de supports numériques classiques comme des ordinateurs, des téléphones portables, des appareils de photographie, des mémoires de stockage, des enregistreurs de vidéosurveillance et des solutions *Global Positioning System* (GPS). Les méthodologies de travail standardisent et encadrent la démarche d'appréhension de ces appareils numériques. Toutefois, l'Internet des objets apporte de nouveaux matériels hétéroclites et non standardisés, communiquant sur une pluralité de protocoles propriétaires ou non. La figure 5.1 présente quelques-uns des protocoles utilisés dans l'IdO triés en fonction de leur débit et de leur dis-

tance de communication. Cette grande variété s'explique par des cas d'usages différents. De plus, ces équipements se confondent dans l'environnement de proximité des personnes et de l'habitat. Ils ne sont pas toujours manifestes et identifiables visuellement. À ces caractéristiques de perception, il convient d'ajouter l'absence de communication permanente, élément discriminatoire dans la distinction des objets physiques. Cette propriété est liée au besoin de limiter la consommation énergétique. Un écosystème connecté est souvent composé de plusieurs dispositifs connectés complémentaires et dispersés. Il comprend de nombreuses dépendances en particulier en matière de politique de gestion de la donnée. Cette situation complexifie la recherche des traces numériques. Elle est sujette à des oublis ou à des erreurs d'appréciation. En l'absence de solutions techniques et méthodologiques adaptés, les enquêteurs ne sont pas en mesure de répondre efficacement à l'appréhension de cet environnement pluriel. La recherche manuelle donne peu de résultats satisfaisants, car ils sont partiels. L'enjeu de cette partie est de donner une image claire et précise des objets numériques présents dans le périmètre limité à une scène de crime. L'environnement connecté est abordé par sa signature électromagnétique et par les spécificités des matériels.

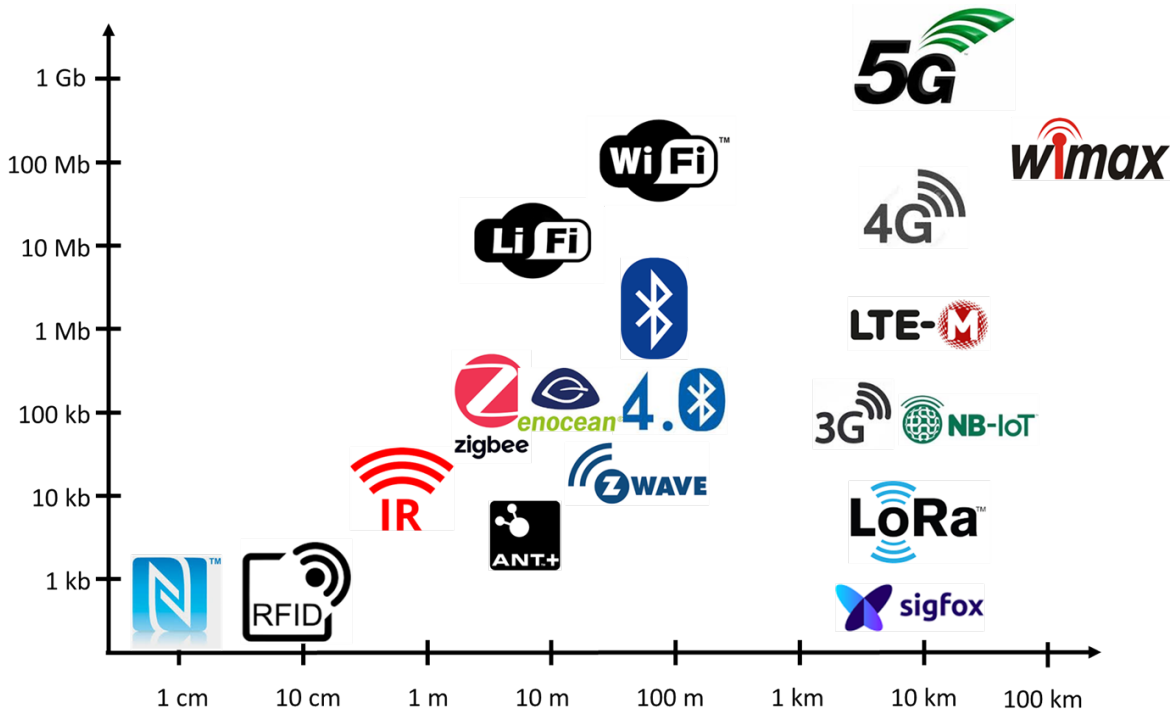


FIGURE 5.1 – Cartographie des protocoles de l'Internet des objets, en fonction du débit et de la portée du signal

5.1.2 Techniques de mesure de la localisation

Il existe déjà de nombreux travaux intéressants sur le concept de la localisation des objets en intérieur et en extérieur. Certaines méthodes de calcul sont liées aux caractéristiques techniques des systèmes telles que l'indicateur d'intensité du signal reçu *Received Signal Strength Indication* (RSSI), la différence de temps *Time Difference Of Arrival* (TDOA), l'angle d'arrivée *Angle of Arrival* (AoA), le déphasage, le nombre de sauts, etc. D'autres procédés de mesure sont dépendants des particularités du réseau, telles que la connectivité. Nos travaux abordent la recherche des dispositifs connectés par le RSSI et le déphasage des signaux. Les autres solutions techniques (TDOA, AoA, etc.) demeurent plus compliquées à mettre en œuvre dans un cadre opérationnel, fortement contraint. Il en est ainsi pour le TDOA qui nécessite une synchronisation de réseau. Par ailleurs, les algorithmes de trilatération et de triangulation sont également préférés dans notre processus de localisation. La trilatération est basée sur la connaissance des distances séparant la cible des différents points de référence et des coordonnées spatiales de ces ancres. La triangulation s'appuie sur l'analyse de l'angle d'incidence du signal émis par l'objet communicant [94].

La force du signal radioélectrique d'un appareil diminue à mesure que la distance de la source augmente. Ce phénomène s'explique par les interactions entre l'onde et le milieu de propagation. Il existe plusieurs modèles pour décrire cette relation : l'équation de Friis [95], le modèle Two Ray Ground Reflection et le modèle Shadowing [96]. L'équation de Friis est basée sur l'hypothèse d'une distribution uniforme de l'énergie dans les sphères concentriques. Elle s'applique à une liaison satellite. Le modèle Two Ray Ground Reflection considère que le signal suit deux voies principales pour atteindre le récepteur : l'une directe et l'autre réfléchie par le sol. Le modèle Shadowing s'intéresse à l'atténuation et à la variation de puissance d'un signal à une distance donnée selon une loi log-normale. Dans ce modèle, l'aire de distribution n'est pas assimilée à un cercle mais à une surface dont les limites varient dans le temps. Certains équipements radioélectriques sont capables de fournir aux couches supérieures une estimation de la puissance reçue en dBm (rapport de puissance en *décibel* (dB)) entre la puissance mesurée et un *milliwatt* (mW)). Sur la base de ces informations, il est possible de déduire la distance parcourue par le signal. La littérature scientifique contient une multitude de solutions pour estimer la position d'un objet communicant à partir du RSSI [97] [98] [99] [100] [101] [102] [103]. Plusieurs articles déclinent ce principe dans le domaine des objets connectés utilisant les protocoles Wi-Fi [104], Bluetooth [105] [106] et ZigBee [107]. Dans Ferreo et al. [108], les auteurs proposent une localisation des émetteurs radioélectriques en position angulaire.

5.1.3 Contraintes spécifiques au terrain

L'infrastructure de l'Internet des objets est constituée localement d'objets connectés hétérogènes et non standardisés. Ces appareils sont petits, de faible puissance et indépendants en énergie. Ils émettent une signature numérique moindre, en une communication intermittente et réduite dans le temps. Les ressources limitées conditionnent la quantité et le type de données à transmettre. À ce constat, il convient d'y ajouter des facteurs de multiplicité et de concentration des émissions dans un périmètre réduit. Ces éléments constituent une contrainte dans leur étude.

Par ailleurs, le RSSI est très sensible à la singularité de l'environnement. Ainsi, une bonne localisation nécessite la caractérisation de la zone cible. Ce travail prospectif consiste à déterminer le type et la taille des matériaux présents. L'étude de la topologie est réalisée à l'aide d'une solution laser ou sonar. Cependant, les spécificités du milieu traversé par les ondes radioélectriques ne sont pas toujours connues. À titre d'illustration, le corps humain atténue les signaux en raison de sa forte teneur en eau [109]. Les modèles généraux offrent la possibilité d'estimer l'atténuation du signal. Ainsi, la puissance reçue est proportionnelle à la distance avec un exposant de perte de trajet (Tableau 5.1). Il est lié à l'environnement traversé [110, 111]. Une composante verticale est ajoutée pour prendre en compte le relief et les niveaux de terrain. Dans un édifice, la pluralité des étages est abordée comme une superposition de plans de mesure. Plusieurs facteurs physiques sont intégrés, en particulier la température et les conditions climatiques.

Environnement	Coefficient de perte
Vide	2
Urbain	2.7 à 3.5
Périurbain	3 à 5
Intérieur (ligne de mire)	1.6 à 1.8

TABLE 5.1 – Coefficient de perte du chemin de propagation

La position relative des antennes conditionne le niveau de puissance détecté. Une étape de profilage est nécessaire en amont de l'acquisition des données. Selon les protocoles détectés, les mesures sont à effectuer sur de longues périodes. Cette remarque s'applique particulièrement dans le cas de protocoles émettant sur de courts intervalles, de façon discontinue. Ainsi, une méthodologie généraliste de la recherche des objets connectés doit être élaborée en fonction des caractéristiques relevées.

5.2 Appréhension de l'environnement connecté par la signature radioélectrique

L'enquêteur judiciaire a besoin d'une méthodologie spécifique dans l'appréhension de l'environnement de l'IdO. Elle est complétée par une palette d'outils polyvalents, répondant aux contraintes opérationnelles des unités territoriales.

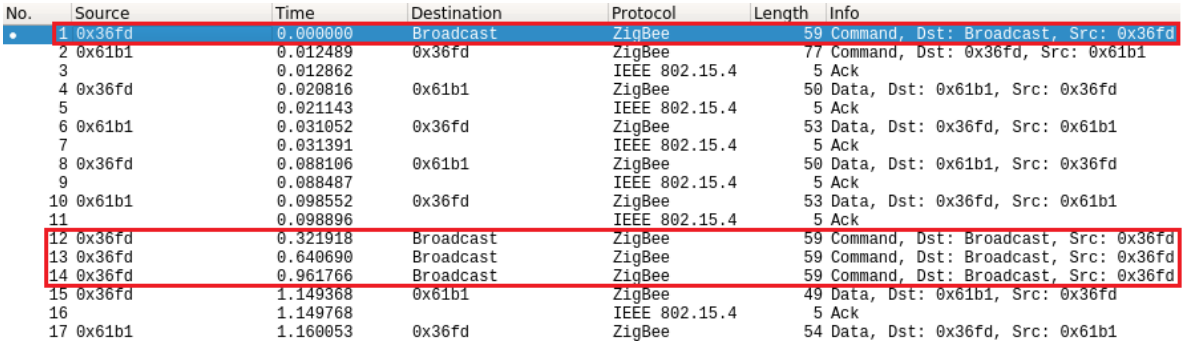
5.2.1 Méthodologie de recherche d'un objet sur une scène de crime

La recherche des objets actifs s'effectue en **quatre étapes** successives. Cette démarche consiste à déterminer les fréquences et les protocoles présents sur un lieu donné, puis, à associer des familles d'appareils par rapprochement et comparaison. Avant toute opération de détection, l'enquêteur doit retirer les matériels de communication qu'il transporte tels que son téléphone portable, sa montre connectée, son ordinateur, les équipements de contrôle - *Procès-Verbal électronique* (PVe) et NéoGend -, etc. Ces appareils sont susceptibles de perturber les mesures, en interagissant avec l'environnement connecté. De la même manière, une discrimination des appareils des premiers intervenants doit être effectuée lors de la phase d'analyse des mesures.

L'**étape 1** consiste à détecter passivement les fréquences et les protocoles exploités. Le balayage fréquentiel est opéré de façon méthodique sur plusieurs niveaux et en plusieurs points de mesure autour de l'environnement à analyser. Cette approche gèle l'atmosphère générale, en identifiant les signaux internes et externes à la zone de mesure. Elle tend également à mesurer l'impact des instruments d'acquisition sur les émissions. Le résultat de cette opération réside en une cartographie générale des communications actives, sur une période donnée. L'**étape 2** comporte l'individualisation des émissions et leur association à des groupes d'objets. Le technicien calcule en plusieurs points le signal radioélectrique sur une fréquence cible, définie à l'étape 1. La mesure est réalisée sur une longue période en cohérence avec l'information recherchée, selon une table des correspondances : fréquence, protocole et durée d'acquisition. Ainsi, avec ces signatures, il détecte la présence des objets. Cette démarche est basée sur la compréhension et la comparaison des caractéristiques techniques des différents protocoles. Leur différenciation se fait dans l'**étape 3**. Les canaux utilisés par les dispositifs cibles sont étudiés un par un. L'influence de l'environnement radioélectrique est prise en compte, ainsi que le comportement des équipements de communication tels que les canaux à vitesse variable, les canaux fixes, la puissance d'émission, etc. Le processus de perception dépend de la sensibilité de la méthode de détection utilisée et la possibilité de différencier la trace de l'arrière-plan.

Cependant à l'issue de cette opération de captation passive, la cartographie des objets demeure incomplète pour plusieurs raisons. Seuls les objets actifs sont détectés. La mesure est limitée dans le temps et donc elle est considérée comme partielle. Par conséquent, il est nécessaire d'utiliser un révélateur des communications. Le *modus operandi* consiste à faire générer un trafic entre les objets. L'impact de cette manœuvre doit être limité et contrôlé afin d'éviter l'altération des objets. Trois solutions sont donc envisagées : générer un ordre de réveil, ajouter un objet connecté au réseau et brouiller une fréquence ou un canal de fréquences précis. Une brève interruption de l'alimentation d'un nœud entraîne une nouvelle synchronisation des objets d'un même réseau. Cependant, cette opération détruit les données stockées dans les mémoires volatiles. Elle modifie la scène de crime et la perte de données utiles. L'**étape 4** consiste en des actions actives et maîtrisées dans le temps et l'espace. Ces manipulations techniques sont effectuées idéalement à l'issue d'une conservation des informations contenues dans les dispositifs observés aux étapes précédentes. Le concept de « commande de mise en éveil » est illustré par le protocole *Simple Service Discovery Protocol* (SSDP), utilisé par certains objets de l'IdO. Une étude approfondie des protocoles de communication doit être réalisée afin de comprendre le principe de la synchronisation des données. Le but ultime est de pouvoir « forcer » l'appariement. Ajouter, brouiller ou déconnecter un périphérique dans une infrastructure de l'IdO génère du trafic radio entre les périphériques. Ainsi, les appareils envoient périodiquement des « messages Hello » pour détecter d'autres membres du réseau. Par exemple, dans notre scénario, le hub ZigBee *Philips* envoie périodiquement des messages en diffusion globale (*broadcast*) comme le montre les captures de trames des figures 5.2 et 5.3. En l'absence de réponse, les liaisons de communication sont coupées ou les objets choisissent de se réorganiser. Par exemple, le protocole de routage pour les réseaux à faible consommation et à perte RPL illustre bien ce concept. Le routage est basé sur la proximité des objets, pour obtenir une communication la plus efficace possible. Toute modification d'un objet du réseau a un impact sur l'organisation du réseau. Un lien d'interdépendance existe entre les objets. Le brouillage peut également être utilisé pour forcer certains protocoles à utiliser des canaux spécifiques.

Les opérations de détection sont rendues plus complexes par la présence de nouveaux protocoles standards. Seuls le type et les entêtes de ces communications sont entendus. Nombre d'entre eux utilisent la diffusion spectrale, la commutation de fréquence à haute vitesse, la modulation du premier signal envoyé en réponse, la communication cryptographique sans possibilité de distinguer deux dispositifs.



No.	Source	Time	Destination	Protocol	Length	Info
1	0x36fd	0.000000	Broadcast	ZigBee	59	Command, Dst: Broadcast, Src: 0x36fd
2	0x61b1	0.012489	0x36fd	ZigBee	77	Command, Dst: 0x36fd, Src: 0x61b1
3		0.012862		IEEE 802.15.4	5	Ack
4	0x36fd	0.020816	0x61b1	ZigBee	50	Data, Dst: 0x61b1, Src: 0x36fd
5		0.021143		IEEE 802.15.4	5	Ack
6	0x61b1	0.031052	0x36fd	ZigBee	53	Data, Dst: 0x36fd, Src: 0x61b1
7		0.031391		IEEE 802.15.4	5	Ack
8	0x36fd	0.088106	0x61b1	ZigBee	50	Data, Dst: 0x61b1, Src: 0x36fd
9		0.088487		IEEE 802.15.4	5	Ack
10	0x61b1	0.098552	0x36fd	ZigBee	53	Data, Dst: 0x36fd, Src: 0x61b1
11		0.098896		IEEE 802.15.4	5	Ack
12	0x36fd	0.321918	Broadcast	ZigBee	59	Command, Dst: Broadcast, Src: 0x36fd
13	0x36fd	0.640690	Broadcast	ZigBee	59	Command, Dst: Broadcast, Src: 0x36fd
14	0x36fd	0.961766	Broadcast	ZigBee	59	Command, Dst: Broadcast, Src: 0x36fd
15	0x36fd	1.149368	0x61b1	ZigBee	49	Data, Dst: 0x61b1, Src: 0x36fd
16		1.149768		IEEE 802.15.4	5	Ack
17	0x61b1	1.160053	0x36fd	ZigBee	54	Data, Dst: 0x36fd, Src: 0x61b1

FIGURE 5.2 – Écoute d'une communication ZigBee du hub Philips avec ZBWireshark faisant apparaître l'envoi périodique de paquets en broadcast pour tenter de joindre des appareils à proximité

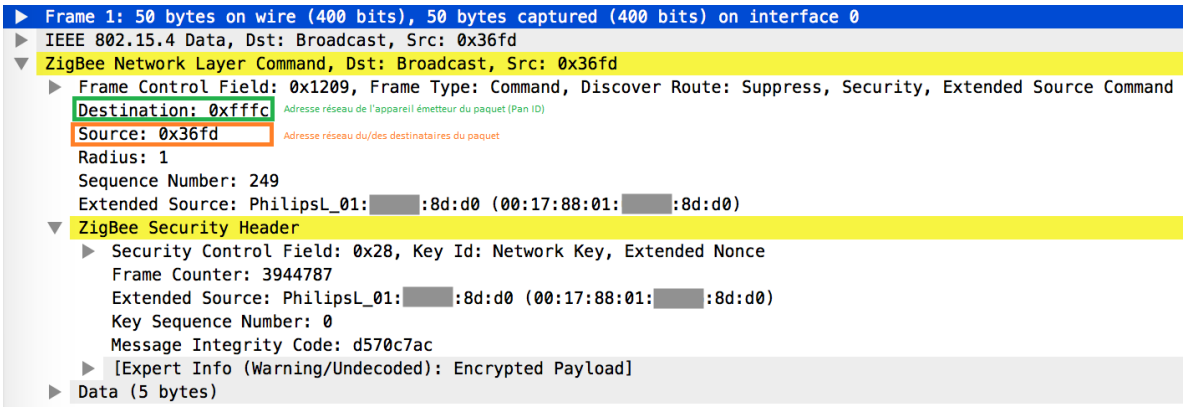


FIGURE 5.3 – Structure du paquet émis par le hub Philips

5.2.2 Reconnaissance et différenciation des objets

À partir des signaux capturés, il est possible d'identifier l'équipement connecté. Les réseaux maillés Bluetooth, Wi-Fi, Li-Fi et ZigBee utilisent le contrôle d'accès au média (MAC). Ce mécanisme d'adressage identifie le fabricant et le produit de manière unique. L'en-tête qui contient la MAC n'est pas chiffrée : 3 premiers octets identifient le constructeur et les 3 derniers octets le matériel. Cette démarche d'identification n'est toutefois pas généralisable à tous les protocoles, comme par exemple avec un protocole de routage *ad hoc* ou avec le protocole *LoRa Wide Area Network* (LoRaWAN). De même, l'adresse MAC reste modifiable ou altérable par une action volontaire de l'utilisateur. Cette situation est peu commune lors d'une intervention judiciaire sur une scène de crime. Elle correspond à un type spécifique de délinquance nécessitant l'intervention d'unités spécialisées. L'écoute des communications offre en outre une discrimination des matériels à partir de leurs propriétés physiques telles que la plage de fréquence de transmission, la modulation, la forme du signal, le temps de réponse,

la puissance du signal, etc. Elle nécessite de coupler cette mesure de signal avec une solution d'apprentissage et une base technique de comparaison.

Il est important de souligner que la capture d'une MAC est considérée dans certains pays comme une violation de la vie privée. Elle constitue une donnée à caractère personnel. Dans un cadre judiciaire, cette information doit être rattachée au lieu de l'intervention des forces de police. La phase de localisation des appareils répond à cette problématique en cloisonnant l'étude des signaux à ceux liés uniquement à la scène de crime.

Plusieurs solutions dans la recherche des objets connectés en fonction de leurs caractéristiques d'émission sont proposées ci-après. Elles répondent à des besoins et des exigences opérationnelles spécifiques.

5.2.3 Outils de recherche forensique par radiofréquence

Il existe plusieurs outils pour appréhender l'environnement électromagnétique : une radio logicielle (ou *Software Defined Radio* (SDR)), un capteur basé sur un protocole unique de communication et un réseau maillé de capteurs. La SDR renvoie les fréquences utilisées. Le récepteur unique offre une cartographie globale de l'environnement sur un protocole donné. Le réseau maillé multi-capteur donne une vision précise et ciblée de l'infrastructure connectée selon plusieurs protocoles et fréquences.

5.2.3.1 Radio logicielle

La radio logicielle est utilisée pour déterminer les émissions présentes dans la zone d'étude. Cet outil balaye l'ensemble du spectre de fréquences. Il détermine celles qui sont utilisées. Selon le résultat obtenu, le technicien détermine rapidement et passivement la présence d'objets actifs. Cette mesure offre une levée de doute sur la scène de crime. La brique logicielle dédiée à la localisation n'est pas développée dans ces travaux de recherche. Ce choix est motivé par des questions de coût de reprogrammation et de déploiement d'une solution SDR. Ainsi, la localisation est réalisée par deux outils dédiés : un seul récepteur ou un réseau maillé multi-capteur. Le choix d'un outil est conditionné par les conditions d'acquisition, l'environnement de mesure, la fréquence et les protocoles trouvés avec la SDR.

5.2.3.2 Récepteur unique

Le récepteur unique est une variante de l'approche SDR basée sur une puce dédiée à un protocole unique. Un capteur mobile balaye la zone d'étude sur une plage de fréquences connue. Il donne une image globale des objets actifs de la même famille de protocole. Les mesures sont donc passives sans interférence, ni injection de charge utile, afin de respecter les

aspects légaux. Ils sont effectués en se déplaçant dans la zone cible. La localisation est basée sur la valeur du RSSI.

Le module de mesure se compose de quatre parties (Figure 5.4) : des antennes à secteur de fréquence fixe, un système de localisation automatique composé d'un GPS, des accéléromètres et des gyroscopes, un système d'exploitation *Operating System* (OS) et une batterie. Dans nos travaux de recherche, les antennes à grande vitesse avec une grande grille de surface ont été retenues.

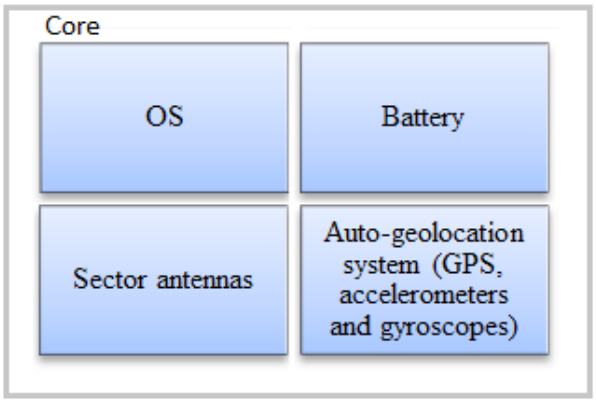


FIGURE 5.4 – Diagramme fonctionnel du récepteur unique

Pour améliorer l'efficacité et réduire la zone cible, la mesure est corrélée à des informations externes provenant des réseaux de communication tels que le système mondial de communications mobiles *Global System for Mobile Communications* (GSM) et le Wi-Fi, en particulier dans la recherche d'un téléphone mobile. L'utilisation de drones contenant le récepteur unique facilite l'usage de cette technologie de mesure pour des espaces vallonnés ou accidentés.

5.2.3.3 Réseau maillé multi-capteur

Le réseau maillé multi-capteur est basé sur l'étude statique d'une zone spécifique à partir de plusieurs capteurs synchronisés. Ils ont la particularité d'appartenir au même réseau maillé. Ils sont contrôlés par une centrale de contrôle et commande. Cette solution donne un positionnement précis des objets connectés en trois dimensions. Elle s'appuie sur le RSSI et le déphasage entre les capteurs.

Le multi-capteur est composé de six parties (Figure 5.5) : des antennes à secteur de fréquence fixe, un système de localisation automatique composé d'un GPS, d'accéléromètres et de gyroscopes, un système d'exploitation (OS), une batterie, un système de communication maillé et un module de synchronisation. Il est accompagné d'un système sans OS dédié à l'étalonnage du capteur. Il capture des informations externes telles que la température, la

pression, l'ouverture et le mouvement. Le module d'antenne est composé de plusieurs antennes observant des secteurs de 120 degrés horizontaux sur 15 degrés verticaux. Les antennes omnidirectionnelles polarisées verticalement par rapport au sol ou de type quart d'onde ont été retenues dans notre étude. Elles ont l'avantage d'être directionnels et de petites dimensions. Afin de saisir la configuration du site, un magnétomètre a été ajouté en complément de la visualisation tridimensionnel (3D).

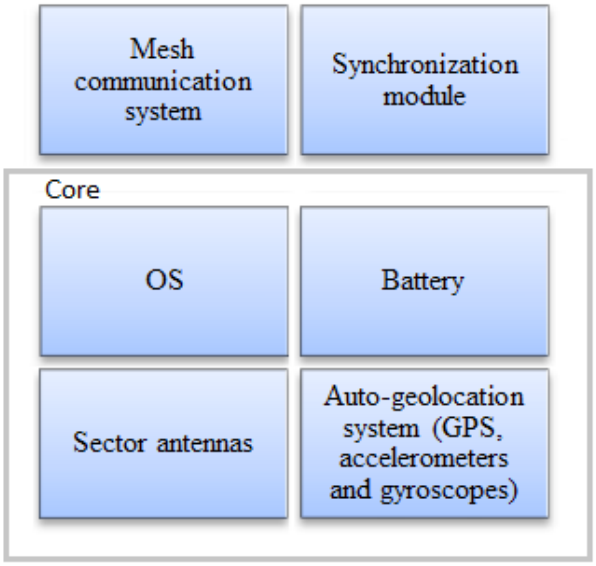


FIGURE 5.5 – Diagramme fonctionnel du multi-capteur

Initialement, les capteurs sont situés aux trois extrémités de la zone d'étude. Pour améliorer la précision des mesures, un capteur est placé en son centre. Un signal est transmis pour déterminer le positionnement des capteurs et les étalonner. La synchronisation des horloges est régulée par radiofréquence. Afin d'éviter la perturbation ou l'altération de l'environnement numérique, l'étalonnage des capteurs du réseau maillé multi-capteur est effectué en mode *promiscuous*. Une analyse spectrale est donc opérée sur une durée suffisante pour déployer des capteurs sur des fréquences non utilisées.

Cette infrastructure de réseau analyse l'ensemble du spectre des bandes de fréquences *Industrielles, Scientifiques et Médicales* (ISM). Les informations recueillies guident le déploiement de nouveaux capteurs configurés sur des protocoles cibles. Ces nouveaux arrivants améliorent la précision de la localisation des objets connectés présents sur la scène d'intervention. Ils facilitent également la recherche d'objets cachés. L'approche recommandée consiste à effectuer une mesure globale de l'activité fréquentielle. Ensuite, elle est déclinée en études concentriques sur des environnements clos. La vitesse et la qualité de la synchronisation entre les capteurs et la continuité des communications impactent significativement la mesure effec-

tuée. Les mesures sont également couplées à des informations provenant de réseaux telles que les données GSM et Wi-Fi.

5.2.3.4 Stratégies d'évaluation

Les critères de la durée d'appréhension de l'environnement ainsi que la connaissance des protocoles utilisés sont les éléments déterminants pour favoriser le choix d'un outil. En effet, le temps d'acquisition des mesures est contraint par des exigences opérationnelles ou juridiques. À titre d'illustration, une perquisition lors d'une mesure de garde à vue est limitée dans le temps. À l'inverse, le temps est moins contraint dans le cadre d'un traitement d'une scène de crime. À ce critère, il convient d'ajouter les conditions de l'acquisition des données.

Le récepteur unique est souvent utilisé dans la recherche d'un objet connecté dont le protocole de communication est connu en amont. Cette remarque se justifie par l'usage d'une chaîne de mesures dédiée (matériel et logiciel). Cet équipement est également très efficace dans le traitement de grandes zones d'étude en particulier par l'usage de drone ou de moyens mobiles. Le réseau maillé multi-capteur est plus complexe à mettre en œuvre en raison du problème de synchronisation entre les capteurs. Cependant, il donne une vision précise et ciblée de l'infrastructure connectée combinant plusieurs protocoles de communication, connus ou non. Cet outil prend en charge des protocoles propriétaires et conventionnels tels que Wi-Fi, ZWave, Bluetooth, ZigBee, XBee, SigFox, LoRa, etc. Il est également utilisé dans la recherche d'objets émettant sur des fréquences interdites d'utilisation. Les deux méthodes sont combinables afin d'optimiser l'étude d'un environnement. Dans ce cas, la démarche d'appréhension se développe selon une approche globale avec le récepteur unique vers une démarche plus ciblée et locale avec le réseau maillé multi-capteur.

5.3 Expérimentation sur une scène de crime

Les solutions d'appréhension de l'environnement connecté sont évaluées au cours de l'exercice d'enquête défini et décrit dans le chapitre 3 de la thèse. Un retour d'expérience complète cette évaluation en tenant compte des contraintes opérationnelles dans sa mise en application en mission.

5.3.1 Description des conditions et paramètres de l'expérience

L'expérimentation a pour objectif d'étudier les performances et l'efficacité des solutions proposées. L'évaluation se fonde sur deux critères de mesure. Le premier est un rapport entre le nombre d'objets détectés par rapport à ceux présents physiquement dans l'appartement. Le second calcule le temps de recherche des dispositifs connectés. L'échantillon est constitué

d'objets connectés domestiques hétérogènes (domotique, santé et assistants virtuels) et de leurs passerelles. Ces équipements communiquent suivant différents protocoles commerciaux (Ethernet, Wi-Fi, Bluetooth, ZigBee, SigFox) et propriétaires (en 915 MHz, pour l'écosystème *Sen.se*).

Une première série de mesures avec une SDR est réalisée pour déterminer les fréquences utilisées. Afin de ne pas altérer la zone d'étude, les mesures sont effectuées uniquement autour de l'appartement, conformément aux opérations réalisées dans un cas réel d'enquête. Une deuxième série de mesures avec le récepteur unique est exécutée sur différentes hauteurs afin d'intégrer la troisième dimension. Enfin, une dernière série est effectuée avec le réseau maillé multi-capteur. Trois capteurs sont positionnés autour de la scène d'expérimentation. Un quatrième capteur est placé en son centre. Les fréquences extérieures au périmètre d'étude sont préalablement discriminées afin de faciliter l'étude. Cette exclusion est basée sur l'étude de l'emplacement des sources émettrices. L'opération d'acquisition des données est répétée plusieurs fois sur plusieurs jours, selon des conditions climatiques identiques, par des équipes différentes, en une approche en boîte noire de l'appartement. Elle est complétée par une recherche manuelle des appareils. La température de l'air ambiante intérieure est de 19 degrés et celle de l'extérieur de 7 degrés en moyenne.

5.3.2 Résultats et discussion

Le rapport de performance est déterminé par le nombre d'objets détectés par rapport à ceux présents. Le taux d'efficacité est estimé à partir de la durée d'acquisition des objets. La précision des mesures de localisation est également étudiée (cf. Tableau 5.2).

Afin d'installer le quatrième capteur du réseau maillé, il est nécessaire d'accéder à l'appartement témoin. Cette opération est inscrite dans notre traçabilité de mesure. Elle génère une communication ZigBee entre le capteur d'ouverture de la porte et la passerelle *Orvibo*. Cet événement est enregistré et détecté par le réseau multi-capteur. Avec un seul récepteur, la signature n'est pas détectée en l'absence de chaîne de mesure dédiée en 915 MHz.

Les outils offrent la possibilité de trouver plus d'objets connectés qu'une approche manuelle, selon un coefficient de 2 dans cette expérimentation. Le *Terraillon Dot* n'est pas détecté par les outils de mesure. La communication Bluetooth n'est active qu'après une synchronisation manuelle des données. Aucune communication n'est détectée entre les lampes et la passerelle *Philips*. L'émission d'un signal est la conséquence d'un événement généré par une action extérieure liée à l'application de contrôle, d'une programmation ou d'une commande vocale par l'assistant *Amazon Echo*. *Sens'it* est basé sur le protocole SigFox. Il fournit un

No	Appareil	Fabricant	Manuel	SDR	Récepteur unique	Réseau maillé	Outils combinés
1	Capteur d'ouverture	Orvibo	N	N	N	D / L	D / L
2	Capteur d'ouverture		N	N	N	N	N
3	Capteur de présence		N	N	N	D / L	D / L
4	Caméra		N	D	D / L	D / L	D / L
5	Passerelle		L	N	N	D / L	D / L
6 - 7	Ampoule	Philips	N	N	N	N	N
8	Passerelle		L	D	D / L	D / L	D / L
9 - 13	Cookie et Mother	Sen.se	N	D	N	D / L	D / L
14	Echo Spot	Amazon	L	D	D / L	D / L	D / L
15	Pi0	Raspberry	L	D	D / L	D / L	D / L
16	IP Camera M136W		N	D	D / L	D / L	D / L
17	WinkHub 2	Wink	L	D	D / L	D / L	D / L
18	Watch 3	Apple	L	D	D / L	D / L	D / L
19	iPhone SE		L	D	D / L	D / L	D / L
20	Dot	Terraillon	N	N	N	N	N
21	Sens'it 2.1	Sens'it	N	N	N	D / L	D / L
22	Bracelet	Heroz	N	N	N	N	N
23	Pèse-personne	Nokia	N	D	D / L	D / L	D / L
Détection (D)			0%	60%	40%	78%	78%
Localisation (L)			30%	0%	40%	78%	78%
Précision moyenne (mètre)			0	0	2.2	1.1	1.1
Durée moyenne (min) :							
- Déploiement			0	1	2	33	36
- Acquisition			45	18	24	54	44
- Désinstallation			0	1	2	33	36
Durée totale (min)			45	20	28	120	116

TABLE 5.2 – Analyse de la performance des outils d'appréhension d'un écosystème connecté (D : Détection, L : Localisation et N : Aucun résultat)

retour ponctuel et limité conformément à la réglementation RC1 [4] : soit 140 messages par jour, avec une durée de transmission de 1,44 secondes pour une taille de 4 octets par message. La détection nécessite donc une longue mesure dans la durée. Le bracelet *Heroz* n'est pas connecté à un réseau, expliquant l'absence de communication de cet objet.

5.4 Retour d'expérience dans le cadre d'une enquête judiciaire

Ce retour d'expérience donne un aperçu de l'usage des outils d'appréhension d'un environnement connecté dans le cadre de contraintes opérationnelles et du terrain. Il vient en complément de l'expérimentation réalisé en laboratoire concernant l'exercice de la scène de crime. Cette étude statistique s'appuie sur 400 missions réparties entre novembre 2015 et 2019. Les dites missions ont été réalisées par des enquêteurs judiciaires du *Centre de Lutte contre les Criminalités Numériques* (C3N), projetés en appui des unités territoriales de la Gendarmerie Nationale française.

5.4.1 Propriétés des environnements de mesure

Le tableau 5.3 montre l'utilisation des outils développés en fonction des caractéristiques de la zone de mesure. Les missions sans équipement se déroulent principalement dans des zones résidentielles pour des environnements urbains et villageois. La méthode dite « manuelle » est limitée en termes d'efficacité dans la recherche d'objets cachés sur une large zone de couverture. L'équipement SDR répond à ce besoin, notamment en campagne ou en forêt. Il autorise également des levés de doutes, en particulier dans le cadre d'une intervention sur site. Le capteur unique est principalement utilisé dans les zones urbaines et périurbaines. Cet outil offre une approche globale de la situation avec l'usage de drones ou de moyens mobiles. Il fournit également une réponse efficace dans la recherche ciblée d'objets connus sur les protocoles Bluetooth et Wi-Fi. Le multi-capteur est utilisé dans des zones hétérogènes. Ce résultat s'explique par la capacité de l'outil à couvrir un plus grand nombre de protocoles et de fréquences différents. Toutefois, son utilisation est plus restrictive d'un point de vue opérationnel, car il

4. RC (SigFox Radio Configuration) : découpage par zone géographique des pays (de RC1 à RC7) autour de paramètres communs : une gamme de fréquence, une puissance rayonnée maximale et des spécificités du frontal radio. RC1 - Liaison montante (fréquence centrale et débit) : 868.13 MHz et 100 bit/s - Liaison descendante (fréquence centrale et débit) : 869.525 MHz et 600 bit/s - *Effective Isotropic Radiated Power* (EIRP) : 16 dBm - Région : Europe, Oman, Afrique du Sud, Iran et Maurice. Le cycle d'utilisation est de 1% du temps par heure (36 secondes). Pour une charge utile de 8 à 12 octets, cela signifie 6 messages par heure, 140 messages par jour.

demande du temps et des connaissances avancées dans sa manipulation. Des solutions mixtes sont utilisées dans les environnements complexes ou mixtes. Elles répondent aux problèmes de ciblage et de couverture d'une large zone.

Caractéristiques	Nombre de missions	Urbain	Périurbain	Campagne	Village	Forêt	Montagne	Zone sensible
Manuel	12	33.33%	8.33%	16.67%	41.67%	0.00%	0.00%	0.00%
SDR	34	14.71%	5.88%	32.35%	11.76%	35.29%	0.00%	0.00%
Récepteur unique	200	37.00%	29.50%	11.00%	22.00%	0.50%	0.00%	0.00%
Réseau maillé	110	15.55%	20.00%	24.55%	16.36%	18.18%	0.00%	6.36%
Outils combinés	44	11.36%	9.09%	6.82%	6.82%	11.36%	18.18%	36.36%

TABLE 5.3 – Usage des outils en fonction des caractéristiques d'environnement

Le tableau 5.4 se concentre sur les caractéristiques géographiques et climatiques rencontrées par les enquêteurs lors des acquisitions des mesures. La couverture d'une zone résidentielle et de ses environs est estimée à environ 2,5 km². Les espaces sans logement représentent en moyenne de 6 à 10 km². La montagne due aux caractéristiques du relief limite la propagation des ondes et les mesures avec nos outils. Les milieux dits sensibles sont constitués de grandes zones protégées de l'urbanisation et de l'activité civile.

	Superficie moyenne couverte (km²)	Température minimale (T°C)	Température maximale (T°C)	Pluie/Neige
Urbain	2.61	-4	38.1	16%
Périurbain	2.34	-3	39.5	22%
Campagne	6.5	-7	38.6	25%
Village	2.8	-2	36.7	20%
Forêt	10.7	-2.9	35.9	22%
Montagne	3.7	-7	35.2	28%
Zone sensible	15	-3	38	23%

TABLE 5.4 – Caractéristiques géographiques et conditions climatiques d'usage

5.4.2 Résultats des mesures

Le tableau 5.5 donne un aperçu du nombre de signaux radio détectés par mission. Selon la spécificité du matériel utilisé, il ressort que le réseau maillé multi-capteur donne des résultats beaucoup plus précis. Le couplage des outils permet une capture maximale de signaux.

	Minimum	Maximum	Moyenne
SDR	2	45	24
Récepteur unique	15	2428	834
Réseau maillé	9	2498	1237
Outils combinés	8	2615	1280

TABLE 5.5 – Nombre de signaux détectés lors d'une mission par l'équipement

Le tableau 5.6 présent le nombre d'objets connectés découverts au cours des différentes missions. Il ressort que l'usage d'outils de mesure révèle plus d'équipements qu'une approche manuelle, allant jusqu'à un rapport de 2 à 4 lors de leur combinaison.

	Minimum	Maximum	Moyenne
Manuel	0	22	6.04
SDR	0	22	5.9
Récepteur unique	0	45	9.6
Réseau maillé	0	60	9.2
Outils combinés	0	80	14.6

TABLE 5.6 – Nombre d'objets trouvés lors d'une mission par équipement

Le tableau 5.7 montre les performances des outils utilisés. Les pourcentages se réfèrent au nombre d'objets trouvés avec les outils en complément d'une approche manuelle. Ils permettent d'évaluer la valeur ajoutée des outils pour le terrain. La granularité apportée par le réseau maillé multi-capteur offre une découverte plus importante d'objets connectés que pour les autres solutions. Nous observons une complémentarité de solutions dans la combinaison des outils. Nous ne sommes pas en mesure de connaître le nombre réel d'objets présents et leur nature dans l'environnement d'étude. Le rapport de performance entre le nombre d'objets trouvés et le nombre réel n'est pas calculé.

	Manuel	SDR	Récepteur unique	Réseau maillé
Manuel	100%	-	-	-
SDR	85%	15%	-	-
Récepteur unique	64%	-	36%	-
Réseau maillé	58%	-	-	42%
Outils combinés	47%	5%	28%	20%

TABLE 5.7 – Proportion d'objets trouvés en mission par le matériel

5.4.3 Discussion

La question du temps de mesure est un facteur primordial et incompressible dans le choix d'une méthode d'acquisition. Ce critère est difficilement quantifiable et généralisable à l'ensemble des cas rencontrés par les enquêteurs. Il se décompose en trois étapes : une phase de déploiement de l'outil de mesure, une phase d'acquisition et de suivi du traitement des données et une phase de désinstallation du dispositif. Le temps d'installation du récepteur unique est de deux minutes en moyenne. Pour un réseau maillé multi-capteur, il est d'environ trente minutes à une heure. Ce facteur temps dépend des caractéristiques du terrain (reliefs, cours d'eau, motifs boisés, bâti, routes, etc.) et de la zone à couvrir. De plus, il évolue en fonction de l'ajout ou du retrait des appareils de mesure au regard des fréquences à étudier. Le temps de retrait est le même que le temps de déploiement. L'acquisition des données dans un simple récepteur est comprise entre dix minutes et douze heures. En moyenne, elle est estimée à une heure. Pour le réseau maillé, elle oscille entre une heure et une semaine de mesure, avec une moyenne de vingt-quatre heures. Le traitement est presque instantané pour un simple récepteur, tandis que pour le réseau maillé, il est de dix minutes. C'est à ce moment que les informations sont collectées, traitées et affichées au travers d'un client léger. L'ensemble de ces temps est cumulatif et est subordonné au choix de la réponse opérationnelle (intervention, perquisition et criminalistique). La donnée recueillie renseigne notre base de connaissance technique utilisée dans l'identification des équipements. La publication des résultats est une tâche non critique et donc pas prioritaire par rapport à la collecte et au traitement. Ce temps parallélisé aux autres opérations n'intervient donc pas dans la durée générale de la mesure.

La précision des mesures est également un facteur important pour faciliter la recherche d'objets dans un environnement complexe. Elle dépend des logiciels et du matériel utilisés. À titre d'exemple, les antennes du simple capteur sont omnidirectionnelles. Celles du réseau maillé multi-capteur sont directionnelles. Cette question est également fortement dépendante des conditions d'expérimentation : le temps et les caractéristiques de l'environnement. Ces facteurs externes sont intégrés dans les calculs de notre solution. Ainsi, le récepteur unique a une précision de l'ordre de 10 mètres. Celle du réseau maillé multi-capteur est d'environ 1

mètre. Les mesures sont plus précises dans des conditions optimales. De très bons résultats sont obtenus avec le protocole Wi-Fi du fait d'un meilleur contrôle de la propagation des ondes. Le récepteur unique donne une précision d'environ 1 mètre et le réseau maillé multi-capteur à environ 80 cm. Inversement pour la ZigBee, la précision du réseau maillé est de 1 à 2 mètres.

Le nombre d'objets connectés rencontrés au cours d'une mission peut aller jusqu'à 80 équipements par campagne de mesure. En l'absence d'outils, le volume d'objets trouvés est divisé par un coefficient allant de 2 à 4. Ainsi, la recherche d'objets connectés est un défi pour les enquêteurs. Elle ne peut pas être entièrement automatisée en fonction des spécificités des lieux et des contraintes opérationnelles. En outre, elle se concentre sur les objets actifs dont les signaux sont clairement visibles. L'étude des protocoles et de la cartographie de l'infrastructure connectée est nécessaire. Elle sert à orienter les recherches et à trouver des signatures cachées. Ce travail est couplé à l'identification et à la classification des dispositifs en fonction de leur rôle dans les communications. Cette démarche fait l'objet de développement dans les chapitres suivants.

5.5 En quelques mots : détection et localisation des équipements connectés sur une scène de crime

L'Internet des objets est composé de dispositifs hétérogènes, difficilement identifiables par les enquêteurs et souvent cachés. Face à ce constat, les acteurs du monde judiciaire ont besoin de solutions techniques pour les aider dans l'appréhension de l'environnement numérique afin de s'acquitter d'un travail fastidieux de recherche manuelle. La signature radioélectrique des équipements connectés répond à ce besoin. Grâce au RSSI et au déphasage, les objets sont détectés et localisés avec précision. En fonction des contraintes opérationnelles et du terrain, les enquêteurs ont besoin d'outils flexibles dans leurs usages, offrant une détection rapide en couvrant de larges espaces - récepteur unique couplé d'une mobilité drone - mais également apportant précision et performance sur tous les types de protocoles standards ou non - réseau maillé multi-capteur- (cf. Figure 5.6).

À l'issue de cette phase d'appréhension, les enquêteurs procèdent à la collecte et à la préservation des preuves. Cette manipulation raisonnée est un défi face à une infrastructure interconnectée avec de nombreuses dépendances et une politique propre de gestion de l'information. Elle doit également répondre aux exigences en matière de protection et de suivi de la trace, tout au long du processus d'acquisition.

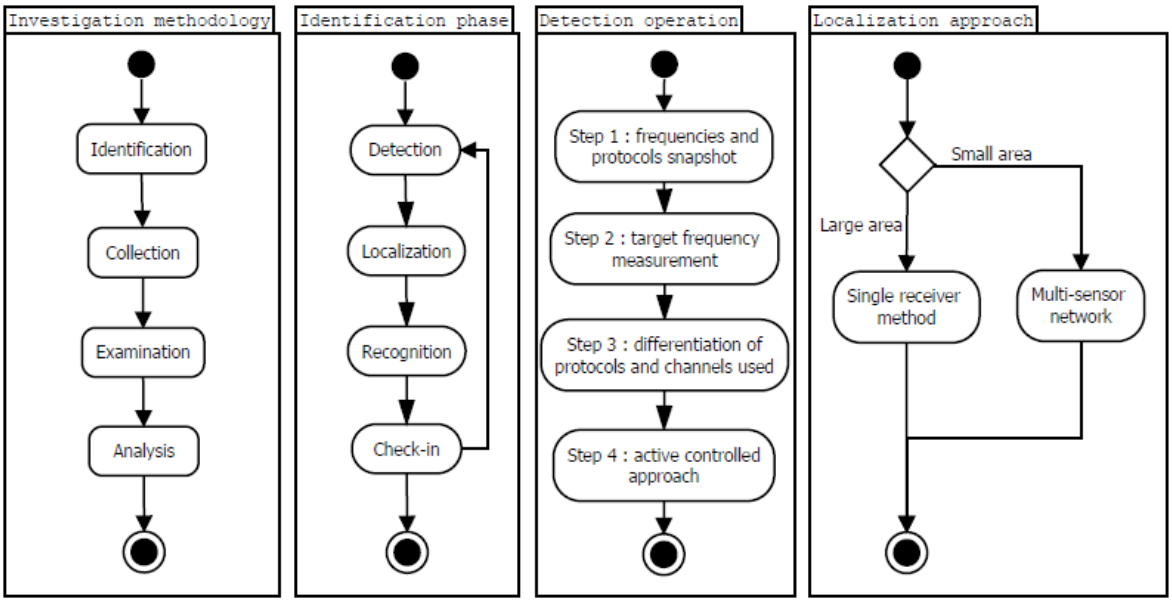


FIGURE 5.6 – Processus d'identification des potentielles sources de preuve dans un environnement local

Chapitre 6

Collecte des sources de traces numériques dans un écosystème connecté

La collecte de traces numériques disséminées dans l'infrastructure connectée est une phase décisive pour le succès de l'enquête judiciaire. Elle consiste à extraire de l'environnement local un ou plusieurs objets connectés sans compromettre l'information stockée pour les placer dans une structure maîtrisée et sécurisée. Cette action offre une conservation optimale des données brutes pour un examen ou une expertise ultérieure.

6.1 Nécessité d'un cadre strict dans la collecte de données numériques

L'opération de collecte répond à des exigences techniques et juridiques liées aux dispositifs étudiés et aux données associées.

6.1.1 Défi de la collecte

La captation des preuves numériques est un défi pour la police technique et scientifique. Plusieurs ouvrages scientifiques énumèrent les difficultés inhérentes à ce sujet. Dans [112], les auteurs les décrivent selon trois caractéristiques : l'architecture, la technologie et les applications. Cet article souligne la complexité d'acquérir des données en raison de l'hétérogénéité des objets à étudier. Cette diversité se traduit par une variété de systèmes d'exploitation et de protocoles de communication plus ou moins uniques [113]. Les caractéristiques techniques

impactent la collecte, en particulier dans le développement d'une approche universelle. Selon [114], le processus d'extraction des preuves dans l'Internet des objets est plus compliqué que dans l'informatique traditionnelle. Ce constat est lié, en autres, aux formats de données traitées, aux protocoles et aux interfaces physiques impliquées. De nombreuses dépendances entre les objets complexifient la donne. Ainsi, un changement dans l'environnement génère une écriture sur le dispositif à collecter, dans les journaux d'événement et le système, voire même une perte d'informations lors de la restructuration de l'infrastructure [115].

Les preuves numériques sont intrinsèquement fragiles. Elles peuvent être altérées, endommagées ou détruites par une mauvaise manipulation ou un examen déficient [116]. Il n'est pas aisé de les copier et de les stocker dans leur état original [117]. Il existe un risque d'arrêt à distance des appareils ou d'écrasement des traces. Il faut donc veiller à documenter et à adapter la méthode de collecte aux contraintes rencontrées. Ces opérations doivent être effectuées en fonction des caractéristiques physiques de l'objet à étudier et des données recherchées. Face à ce constat, une question se pose : quelles mesures de protection doivent être établies pour garantir la non-altération et la conservation optimale des données stockées dans l'infrastructure de l'Internet des objets ? Localement, le challenge est de transférer un dispositif ou un écosystème interconnecté de son environnement naturel vers une nouvelle zone de confinement contrôlée, en évitant le plus possible d'endommager le conteneur et son contenu.

6.1.2 État de l'art de la collecte

La communauté scientifique propose différentes solutions pour gérer les preuves électroniques que ce soit dans l'extraction de données dans l'infrastructure ou la saisie des dispositifs numériques. Néanmoins, le succès de l'exploitation est conditionné par des contraintes opérationnelles et juridiques.

6.1.2.1 Revue de la littérature

Il existe plusieurs stratégies dans la collecte de données au sein d'un environnement connecté. Certains travaux se basent sur l'élaboration d'une interface dédiée, accédant à l'infrastructure existante. Dans [73], les auteurs proposent le modèle *Forensics Aware IoT* (FAIoT). Il s'agit d'une sorte de dépôt central de preuves fiables. Á partir d'une interface de programmation (API) générique, les enquêteurs font remonter les données des plates-formes en ligne. Cette approche nécessite une collaboration forte entre les entreprises privées gestionnaires de l'infrastructure et les forces de police afin de développer un canal d'entrée privilégié au système. Á cette coopération technique, il est opportun d'ajouter les aspects juridiques de la chose, en particulier sur le droit d'accès. Par ailleurs, il existe un biais inhérent au traitement des données. Seules les données synchronisées localement avec les plates-formes sont

accessibles par l'intermédiaire de cette interface. Dans [72], les auteurs proposent de recueillir les données locales en se greffant sur le trafic réseau d'une maison intelligente. Cette solution nécessite une bonne connaissance de l'infrastructure connectée et de son accessibilité. Elle n'est également pas universelle à l'Internet des objets.

L'Internet des objets est une structure plurielle et étendue par nature alliant des éléments physiques au virtuel. Le succès de la collecte est conditionné par un découpage stratégique de l'environnement. Dans [70, 4], les auteurs le divisent en trois domaines d'étude : un environnement local composé d'objets connectés et de leurs passerelles (**zone 1**), un espace ouvert avec les plates-formes Cloud (**zone 2**) et les interfaces de services (**zone 3**). Cette approche est pertinente dans le domaine d'une démarche médico-légale (Figure 6.1). Elle structure le processus d'appréhension de l'infrastructure connectée sur la base des connaissances et des solutions existantes. La zone 1 est constituée de dispositifs hétérogènes et connectés. Elle est unique dans sa configuration, sa topologie et les objets qui la composent. Elle est traitée selon une logique de petits systèmes électroniques domotiques, mobiles ou embarqués. Leur collecte est déterminée par leurs caractéristiques techniques, leurs rôles et leurs dépendances à l'infrastructure locale. La collecte des données de la zone 2 est opérée auprès des opérateurs de l'Internet des objets par des réquisitions judiciaires. Cette action nécessite de cibler la donnée recherchée et d'avoir au préalable clairement identifié les objets présents sur la scène afin d'opérer avec efficacité. Elle est également accessible, dans certaines conditions, par une perquisition « *en ligne* ». La zone 3 est constituée des applications mobiles, des automates de contrôle ou des portails web. La collecte des données passent par l'appréhension des supports physiques d'usage : une interface homme-machine (IHM) dans un boitier de contrôle-commande, un téléphone portable, un ordinateur, etc. Il est à noter que ces appareils jouent dans certaines conditions un rôle déterminant dans les zones 1 et 3. Par exemple, un téléphone portable est en mesure de faire office de passerelle sur un réseau partagé, tout en contenant un applicatif de gestion.

Le processus de collecte des dispositifs physiques est un phénomène irréversible. La surface à appréhender est modifiée par cette opération. Il est donc primordial d'établir un protocole d'échantillonnage rigoureux et raisonné [118, 119]. L'assurance qualité analytique de la scène de crime au laboratoire doit être mise en œuvre [120]. La démarche ne repose plus uniquement sur des considérations juridiques. Les manipulations techniques doivent être effectuées sur les scellés, sur la traçabilité des opérations et sur la continuité des preuves. L'usage de bouclier (*shield*) ou cage Faraday doit être généralisé lors de la saisie ou de l'acquisition afin d'empêcher tout autre trafic radio vers l'appareil. Il est également opportun que l'enquêteur et/ou le laborantin doivent prendre toutes dispositions élémentaires pour ne pas interférer avec l'équipement à étudier. Il existe des approches internationales pour fournir des conseils sur la manière de gérer les preuves électroniques. Le « Guide des preuves électroniques » du

Conseil de l'Europe [121] fournit un cadre aux autorités répressives et judiciaires des pays qui cherchent à établir ou à améliorer leurs propres directives pour l'identification et le traitement des preuves électroniques.

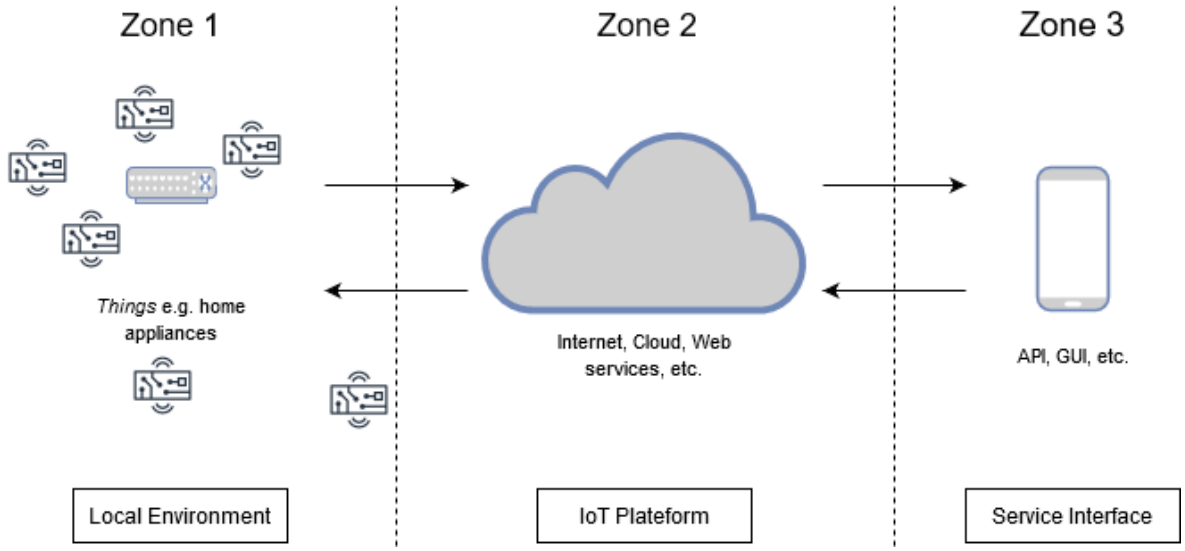


FIGURE 6.1 – Découpage par zone de l'infrastructure Internet des objets

6.1.2.2 Contraintes spécifiques

Souvent les appareils ne sont pas éteints pour préserver les données stockées dans des mémoires volatiles, comme le suggère [70]. L'enquêteur peut alors décider soit d'acquérir les données directement par les méthodes de *live forensic* [122, 123, 124, 125] ou soit procéder à une extraction ultérieure justifiée par le risque d'une altération grave du système ou l'absence d'outils à sa disposition. Dans le deuxième cas, il est nécessaire de maintenir le dispositif dans en l'état pour un traitement ultérieur. L'appareil est éteint uniquement lorsque les données perdues ne présentent pas d'intérêt ou n'affectent pas le reste de l'enquête. Les processus d'extraction des données dans un équipement électronique sont détaillés dans le chapitre 7.

Il arrive que les données soient dispersées sur les appareils du réseau ou sur des services connexes [126]. Cette réflexion fait référence aux dépendances entre les dispositifs et à la politique de gestion de la donnée. Ainsi, il est nécessaire pour l'enquêteur de maîtriser la topologie du réseau. Une phase prospective aborde l'étude du chemin parcouru par les données et la cohérence dans le système. Elle sera affinée au cours de la phase d'analyse de la donnée.

Les limites juridiques de l'enquête constituent un autre défi [75, 70]. Les données voyagent entre plusieurs dispositifs ou services en ligne. Il est souvent difficile de récupérer les informations lorsqu'elles se trouvent sur des serveurs situés dans un pays tiers. En l'absence de

coopération ou d'accord entre les pays et les entreprises privées, la démarche est rendu quasi impossible légalement.

En réponse à ces contraintes techniques, opérationnelles et juridiques, un cadre méthodologique doit être défini pour faciliter la récupération des données et l'appréhension des médias composant l'Internet des objets.

6.2 Méthodologie de la collecte

La collecte d'un écosystème connecté se structure en plusieurs opérations techniques. L'action principale consiste à isoler l'environnement local des interactions avec le monde extérieur et à définir les rôles ainsi que les interdépendances entre les dispositifs. Compte tenu de la topologie des réseaux, plusieurs stratégies sont envisagées conditionnant la saisie et le placement sous scellé.

Dans cette partie, le processus de manipulation des dispositifs connectés, ancrés dans un écosystème local fortement contraint, est étudié. Il est caractérisé par de nombreuses interactions entre les objets. Ainsi, l'étude des zones 1 et 3 contenant les données accessibles techniquement aux enquêteurs sans l'intervention d'un tiers est privilégiée. La zone 2 est abordée au travers de la réquisition judiciaire ou de la perquisition « en ligne ». Elle nécessite néanmoins une compréhension fine du fonctionnement nominal des dispositifs présents en zones 1 et 3 et de la donnée échangée en interzone.

6.2.1 Appréhension de l'environnement

En fonction des contraintes opérationnelles et des éléments recherchés, l'environnement local est traité dans sa globalité ou de façon ciblé par l'enquêteur. Cependant, le processus d'acquisition demeure globalement identique.

6.2.1.1 Approche globale

Cette méthodologie se concentre sur l'examen de la scène de crime dans son ensemble. Elle se compose de trois étapes successives (cf. Figure 6.2). La première étape est l'identification des équipements locaux afin d'obtenir une cartographie des interconnexions et des dépendances. La deuxième étape consiste à déterminer les caractéristiques techniques des équipements et leur fonctionnement nominal. La troisième étape vise à isoler les différents réseaux composant l'environnement local afin de limiter les interactions, les fuites de données et de faciliter l'extraction de tous les équipements.

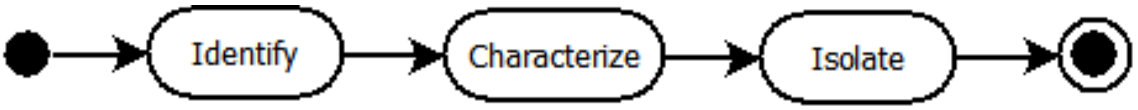


FIGURE 6.2 – Méthodologie d'examen d'une scène locale contenant des dispositifs connectés

Tout d'abord, le technicien cherche à identifier les différentes sources de preuve présentes sur le périmètre cible. Ce processus (détection, localisation, reconnaissance des objets et recoupement de l'information) est initié dès la prise compte de la scène de crime (cf. Figure 5.6 du chapitre 5). Ensuite, il détermine le rôle de ces objets dans l'infrastructure locale. Il existe deux catégories d'objets connectés. La première classe comprend les appareils qui communiquent directement avec le monde extérieur. La seconde contient les objets qui dépendent d'une passerelle pour communiquer. Ces dernières se définissent en fonction des services fournis. Le premier groupe délivre un accès à l'Internet ou à des services externes de la zone 2. Le second assure le lien entre les protocoles spécifiques aux objets et les autres protocoles du réseau. Elles constituent les nœuds de réseau de la zone 1. Des solutions hybrides sont également présentes. Cette articulation est déduite de l'étude des communications, des protocoles et des familles de dispositifs présents. Les objets basés sur des réseaux de courte portée (Bluetooth, ZigBee, Wi-Fi, etc.) communiquent avec l'extérieur via une passerelle locale. Les objets basés sur des réseaux à longue portée (SigFox, LoRa, etc.) utilisent une passerelle publique externe gérée par des opérateurs privés. Sur la base de ces informations, une carte des différents réseaux et l'arborescence générale est dessinée. Les branches symbolisent les communications. La base du tronc symbolise la passerelle principale vers l'Internet. Les passerelles spécifiques aux différents protocoles constituent les nœuds des ramifications. Les objets connectés sont comparables aux feuilles de l'arbre. Cette représentation vise à déterminer les dépendances entre les appareils connectés et à apprécier leurs rôles.

Ensuite, le technicien cherche à identifier et à classer les équipements locaux en fonction de caractéristiques techniques liées au type de mémoire - volatile ou statique - et aux dépendances au sein du réseau. L'objectif est de comprendre comment les données sont synchronisées et le fonctionnement nominal des objets. La synchronisation est automatique, semi-automatique ou manuelle. Le type de données échangées, la politique de gestion des données et leur position dans l'infrastructure sont également des éléments clefs pour déchiffrer l'organisation du réseau. L'identification des objets et leurs modes de synchronisation passent par la base de données techniques de référence. Elle est alimentée lors des retours d'expériences opérationnelles et de la connaissance des familles d'équipements rencontrés.

Troisièmement, le technicien cherche à désagréger et à isoler certaines parties de l'environnement local. Cette approche est abordée du général au spécifique. Les infrastructures locales

sont isolées du monde extérieur en rompant les liens entre les différentes zones (1, 2 et 3). La cartographie de l'infrastructure donne les points d'interaction avec la zone 2. Concrètement, le technicien déconnecte physiquement la communication Ethernet filaire. Il retire les cartes *Subscriber Identity/identification Module* (SIM) ou utilise une solution de brouillage des émissions extérieures. L'interférence est une mesure limitée dans le temps. Elle n'est appliquée que lorsqu'il n'y a pas d'autre moyen d'interrompre la communication. Cependant, l'usage de cette technicité nécessite une étude d'impact sur le périmètre cible. Subséquemment, le technicien traite les différents réseaux de l'environnement local, identifiés lors de la cartographie. Les normes de réseau utilisées sont classables en trois topologies de réseau : point à point, maillage et étoile (cf. Figure 6.3).

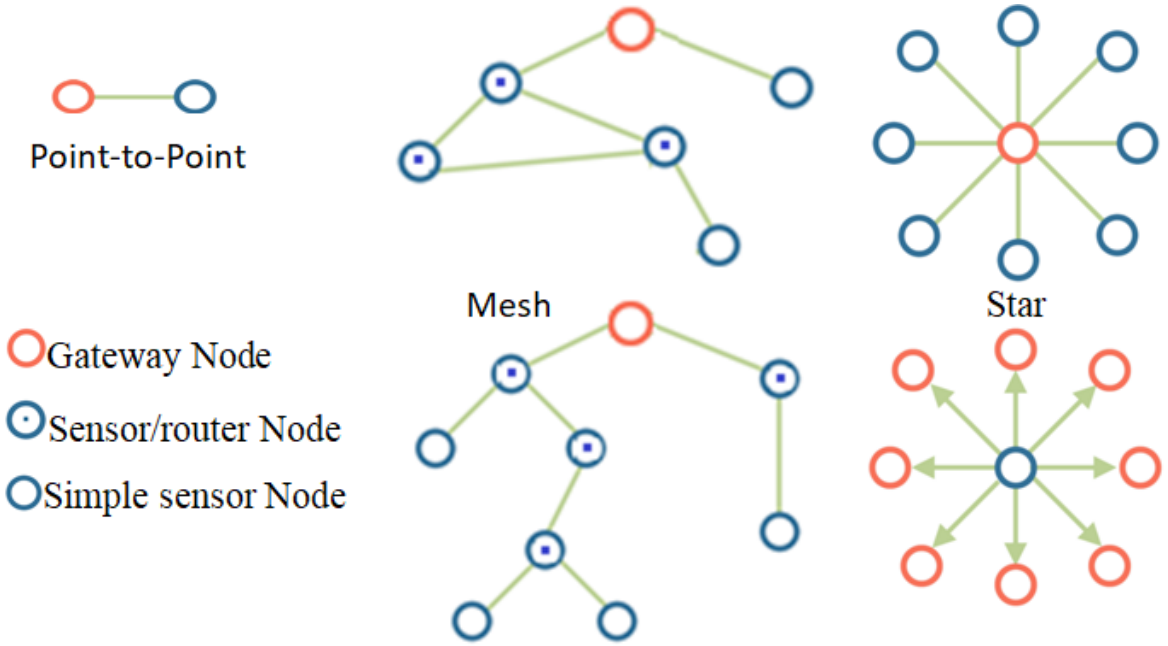


FIGURE 6.3 – Topologies réseau de l'Internet des objets

Le **réseau point à point** établit une connexion directe entre l'objet connecté et sa passerelle. L'équipement accède à l'Internet ou à un autre réseau au travers d'une passerelle dédiée. Un exemple de ce type de réseau est une connexion Bluetooth entre un téléphone mobile et une montre connectée. La rupture de la liaison radio isole la passerelle et l'objet du réseau. Cependant, une modification de la connexion avec une dissociation ou l'appairage d'une nouvelle connexion engendre une perte des données. Cet élément a été identifié lors de l'étude de la montre *Apple Watch*.

Le **réseau en étoile** est constitué d'un nœud central auquel tous les autres nœuds du réseau sont connectés. Cet embranchement principal sert de point de connexion à tous les autres nœuds. Les nœuds périphériques communiquent entre eux uniquement par le biais

d'une passerelle. Cette topologie de réseau s'illustre facilement avec une connexion Wi-Fi d'une habitation. La box est le lien avec le monde extérieur. Cette architecture facilite l'ajout ou la suppression de nœuds, sans impacter le réseau. Toute l'intelligence du réseau est concentrée sur un seul nœud. Cette approche concentrique facilite la gestion du réseau. Ainsi, les boucles périphériques sont à même d'être traitées indépendamment du réseau. En supprimant directement le nœud central, les objets perdent leur connectivité. La structure locale est isolée de toute interaction avec l'extérieur. Cependant, elle est toujours en mesure d'échanger et de stocker des données en interne. Certains réseaux se structurent selon une logique en étoile inversée. L'objet devient le centre du réseau, connecté vers l'extérieur par le biais de plusieurs passerelles périphériques. C'est notamment le cas des objets basés sur les protocoles SigFox et LoRaWAN. Dans ce cas, la solution recommandée est d'isoler l'objet du réseau.

Un **réseau maillé** se compose d'une passerelle et d'objets connectés, dont certains comportent des capacités de routage. Ainsi, un objet est connecté à un ou plusieurs autres objets, agissant comme des nœuds dans le même réseau. La passerelle ouvre un accès vers le monde extérieur. Grâce à ce maillage, les données sont potentiellement relayées par plusieurs nœuds avant d'atteindre leur destination. Ce concept s'appelle un itinéraire. Au fil du temps, les nœuds établissent de nouvelles routes en fonction de leurs états de fonctionnement et des caractéristiques physiques du support. Dans certains cas, cette architecture est hiérarchique. Le nœud parent est le maître du réseau, appelé « arbre des clusters ». Cette structuration est utilisée en domotique selon une construction en relais. Elle compense les problèmes de distance ou de bruit et la présence d'obstacles. Par exemple, le réseau de liaisons intelligentes de la société Enedis est construit sur ce modèle [127]. Pour un réseau maillé hiérarchique, les objets connectés sont isolés du réseau en déconnectant les routeurs. Cependant, cette opération de déstructuration doit partir des extrémités des branches jusqu'au cœur du réseau pour limiter l'écriture dans les logs et une perte de données. Pour un réseau maillé classique, différentes boucles fermées et ouvertes composent le réseau. Les différentes boucles sont isolées entre eux, formant des sous-réseaux indépendants. Les boucles ouvertes sont traitées selon une approche hiérarchisée. Les boucles fermées sont considérées comme un ensemble connecté. Elles ne peuvent pas être déstructurées, sans altération.

L'étude de la topologie du réseau donne une première lecture de l'environnement afin de définir une stratégie de collecte opérante et efficace. Selon les protocoles en vigueur, différentes mesures de confinement sont établies. Elles tiennent compte des dépendances et des hiérarchisations.

6.2.1.2 Cas particulier de la recherche ciblée

La recherche ciblée est un cas particulier de l'approche globale d'un environnement local. Elle est appliquée lors du traitement d'un objet spécifique, en faisant abstraction des autres dispositifs. L'objectif est de promouvoir l'efficacité et la rapidité opérationnelle. Ainsi, cette approche comprend toutes les étapes décrites ci-dessus. Cependant, elle se concentre directement sur l'écosystème de l'objet cible et ses dépendances. Elle fait fi des autres structures. Une cartographie du réseau est réalisée. Elle donne une vue d'ensemble de l'environnement et des équipements connectés présents. Sur la base de ces informations, le technicien dispose de l'arbre des dépendances entre les appareils. Au lieu d'étudier toutes les branches, il cible le réseau qui l'intéresse. Il étudie uniquement les caractéristiques de l'objet cible et des dispositifs qui y sont rattachés. Cette opération nécessite le traitement de tous les équipements d'une même branche.

Après avoir examiné le processus de prise en charge des équipements connectés de l'infrastructure locale, le technicien s'intéresse à leur extraction et à leur conditionnement. Cette opération judiciaire comprend le placement sous-scellé des contenants et de la préservation du contenu face à toute agression extérieure ou intérieure : le temps, les mécanismes d'autodestruction, les contraintes naturelles, les interactions physiques ou électromagnétiques, etc.

6.2.2 Sceller et conditionner la preuve numérique

Après avoir appréhendé l'environnement local et ses équipements, le technicien opère leurs extractions et leurs conditionnements. En fonction des caractéristiques techniques et des capacités techniques disponibles, les données sont extraites directement des médias et du réseau ou ultérieurement dans un environnement contrôle tel qu'un laboratoire criminalistique.

6.2.2.1 Saisir et placer des scellés légaux

La saisie et l'imposition de scellés légaux sont un acte de police judiciaire. Elles consistent à mettre un objet ou un document à la disposition de la justice pour qu'il soit exploité en vue de la manifestation de la vérité [5]. Les saisies sont effectuées à la suite de la remise spontanée d'un objet ou à la suite d'une perquisition judiciaire. Tout comme les objets physiques, les données numériques stockées sur des supports de mémoire sont susceptibles d'être saisies. L'enquêteur scelle soit le support physique contenant les données, soit une copie des données. Avant la collecte, des mesures sont prises pour préserver les traces biologiques, telles que l'ADN ou les empreintes digitales.

5. Les saisies et leurs mises en oeuvre sont prévues par les articles 54, 56, 76 et 97 du CPP.

Le sceau doit garantir l'intégrité des données présentes dans l'appareil. Il protège le contenu de toute interaction physique ou numérique avec l'extérieur. Pour les preuves numériques, il protège l'information contre l'exposition aux champs électromagnétiques. Ainsi, il doit être apposé en fonction des caractéristiques et l'état de l'objet. Si le support ne peut pas être désactivé ou doit rester actif, l'objet est conditionné dans une cage de Faraday. Cette protection doit intégrer une alimentation électrique continue de type batterie indépendante afin de garantir le bon fonctionnement. Cette action augmente artificiellement la durée de vie du contenant. Cependant, elle ne garantit pas l'altération avec le temps des données contenues. Le temps de fonctionnement est estimé en fonction de la capacité de l'alimentation électrique et des tâches en court de traitement. Cette information doit être affichée bien en évidence sur le sceau afin d'être supervisée. Si le support peut être éteint, l'objet est conditionné de telle sorte qu'il ne peut être allumé. Le sceau doit être conforme aux dispositions énoncées dans les traités sur la police scientifique numérique [128, 57, 129].

6.2.2.2 Conditionner la preuve

La scène de crime est constituée d'objets et de passerelles connectés entre eux, avec plus ou moins de dépendances. La figure 6.4 reprend les différentes opérations exécutées pour obtenir un conditionnement efficace et efficient. Après avoir vérifié leur état de fonctionnement - marche ou arrêt-, le dispositif éteint est placé directement sous scellé. Un appareil allumé sans dépendance est déconnecté et isolé du réseau. En l'absence de mémoire, il est placé sous scellé à l'état éteint. En présence d'une mémoire, il est traité directement sur le terrain ou en laboratoire, en fonction de l'état des connaissances et des outils d'extraction disponibles. De nombreuses passerelles disposent de ports de communication ouverts avec des services tels qu'un Telnet ou une API dédiée. Ces informations sont recueillies avec l'outil d'identification et de pré-analyse (cf. chapitre 4). Ainsi, une exploitation de ces services offre la possibilité de procéder à une extraction de leurs données (cf. chapitre 7). L'acquisition locale est toujours préférable que ce soit sur le réseau ou sur le dispositif à traiter. Certains objets connectés contiennent également des supports externes de stockage. Une étude du marché de la montre connectée, réalisée au cours des travaux de la thèse, met en évidence la présence de stockage et de cartes SIM dans un tiers des produits grands publics. Les méthodes classiques d'acquisition médico-légale sont applicables à ce type de support électronique. Une vérification des données extraites est nécessaire pour établir leur intégrité à des fins légales. Une comparaison des hachages de la source primaire de données et des données acquises est effectuée. Toutefois, cette vérification n'est pas exécutée dans le cas d'une acquisition directe [92] [130]. En-effet, la source a été modifiée lors de l'extraction.

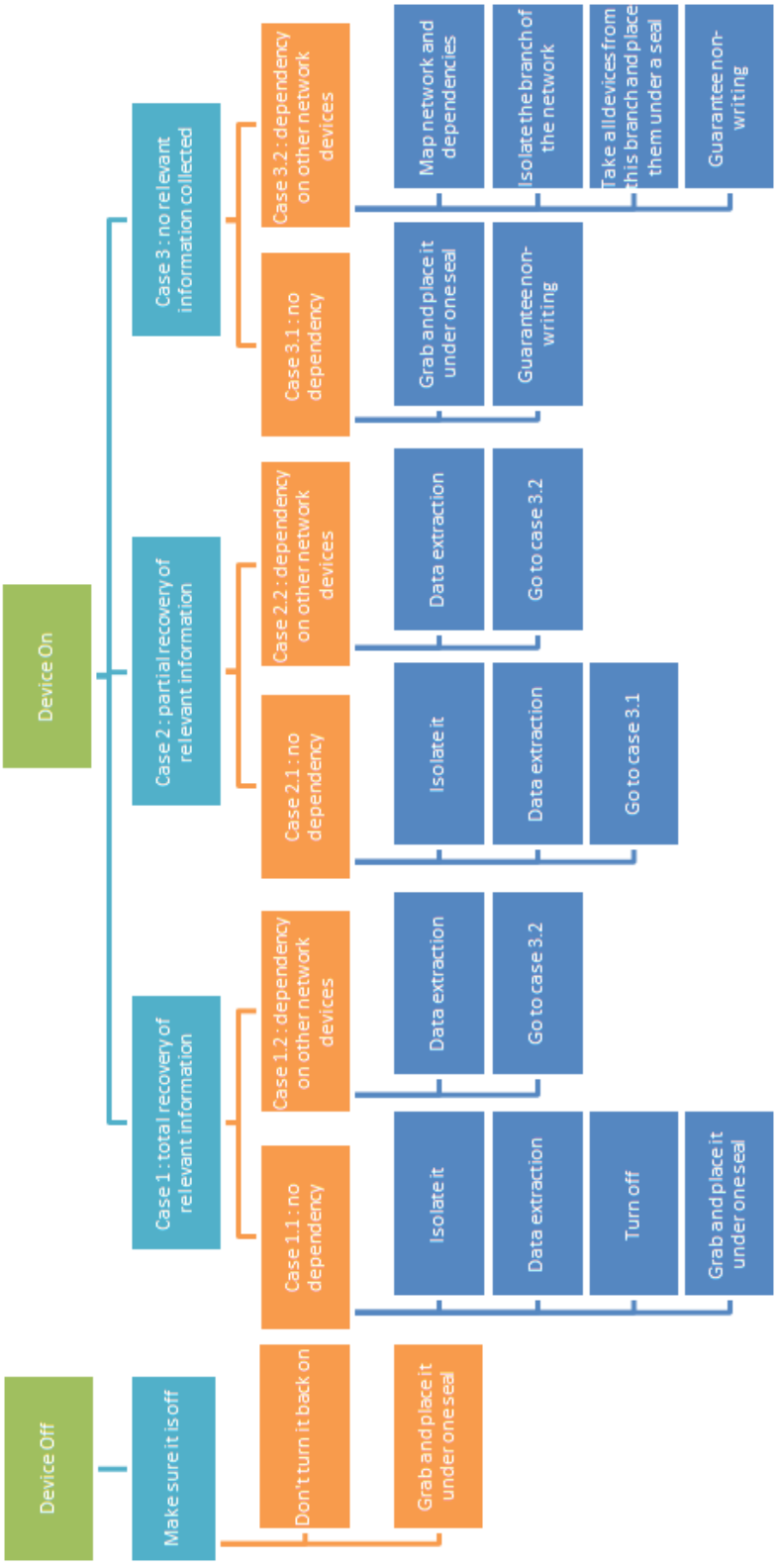


FIGURE 6.4 – Processus de collecte de dispositifs IdO

Lorsqu'il n'est pas possible d'extraire toutes les données pertinentes pour l'enquête, l'équipement est placé dans un état aussi proche que possible de son état d'entrée. Le « mode avion » est activé ainsi que l'arrêt des communications ou d'opérations d'écriture. Cependant, il est à noter que dans certains cas cette action est sujette au lancement de routines de classement des données dans certains supports mobiles. Dans ce cas-là, l'extraction de la carte SIM est privilégiée afin de limiter l'écriture en base. Chaque modification ou opération est mentionnée dans le rapport de traitement de l'objet. À l'issue, le dispositif isolé est placé dans un scellé unique. Pour les objets contenant de fortes dépendances, il est nécessaire de procéder à un scellé commun. En-effet, ils sont considérés comme un tout connecté. Une déstructuration serait source de perte ou d'altération des informations. Toutefois, une étude d'impact doit être réalisée au préalable en déterminant les effets des manipulations à opérer pour un scellé commun ou une multiplicité de scellés uniques. Cette démarche analytique s'appuie sur les retours d'expérience et les connaissances des techniciens provenant d'évaluations réalisées tout au long du processus d'acquisition des équipements (initial, intermédiaire, final et ex-post).

Ainsi, les dépendances entre les objets et l'accessibilité des données avec les outils médico-légaux motivent le choix d'une méthode d'acquisition des équipements et des données sur le terrain.

6.3 Méthodologie de la collecte à l'épreuve d'une scène de crime

La méthodologie de collecte proposée est évaluée au regard de l'exercice d'enquête défini dans le chapitre 4. Une étude d'impact opérationnel est menée.

6.3.1 Analyse de l'environnement local

L'infrastructure locale se compose de quatre réseaux connectés à l'Internet (cf. Figure 6.5) : un réseau structuré autour de l'*iPhone*, un réseau structuré autour du *WinkHub*, un réseau *Heroz* et un réseau *Sens'it*. Chaque entité est traitée indépendamment. Il est à noter la présence d'une connexion entre les réseaux *iPhone* et *WinkHub*.

Le premier réseau contient trois objets connectés : une *Apple Watch*, un *Terraillon Dot* et un pèse personne *Nokia*. La montre intelligente et l'analyseur de sommeil sont connectés au téléphone en Bluetooth. Le pèse personne communique en Bluetooth (ou en Wi-Fi). L'*iPhone* fait office de passerelle. Il fournit un accès Internet aux objets connectés. En déconnectant le *smartphone* du réseau extérieur (GSM, 4G et Wi-Fi), le groupe d'appareils est isolé de tout

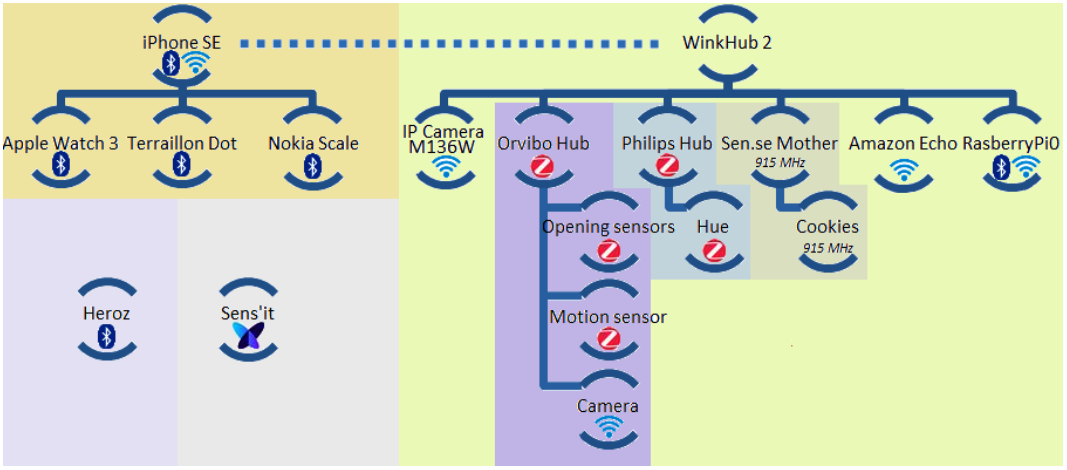


FIGURE 6.5 – Cartographie globale de l'environnement IdO

autre réseau des zones 1 et 2. Ainsi, la carte SIM est extraite du *smartphone*. Les fonctionnalités Wi-Fi et Bluetooth sont désactivées.

Le deuxième réseau contient trois objets connectés : une caméra *IP M136W*, une *Amazon Echo* et un *Raspberry Pi*. Il intègre trois environnements indépendants : *Orvibo*, *Sen.se* et *Philips*. La passerelle *WinkHub* fournit un accès Internet aux différents objets et aux trois environnements connectés. Le lien entre le premier et le second réseau est rompu lorsque la connexion Wi-Fi de l'*iPhone* est coupée. En déconnectant le câble Ethernet du *WinkHub*, le second réseau est isolé du monde extérieur.

Les troisième et quatrième réseaux ne sont constitués localement que de deux objets connectés indépendants : *Heroz* et *Sens'it*. *Heroz* utilise le Bluetooth pour communiquer. Cependant, dans le scénario, il n'est appairé à aucune passerelle. *Sens'it* utilise le protocole SigFox pour communiquer. Il échange directement avec le monde extérieur par le biais de passerelles externes privées, présentes hors de la scène de crime.

Cette première analyse donne quatre réseaux indépendants déconnectés de l'Internet (zone 2). En rompant les liens de radiocommunication, la migration des informations vers des plates-formes externes n'est plus possible. Cette opération garantie également l'absence d'interactions entre ces réseaux.

6.3.2 Caractérisation des équipements connectés

L'étude du fonctionnement nominal des objets et de leurs passerelles aide à déterminer les environnements contenant les données utiles et leur emplacement dans l'infrastructure. Cette analyse s'appuie sur une base technique de données des équipements, renseignée par les différents cas rencontrés et les recherches sur l'identification (cf. chapitres 4 et 8). Les

dépendances entre les objets et les écosystèmes connectés sont également étudiées. Elles sont susceptibles de faire apparaître des liens « cachés », notion également abordée lors de l'analyse (cf. chapitre 7).

Certains objets utilisent la synchronisation automatique de leurs données avec le réseau. Ainsi, ils stockent localement peu d'informations utiles liées à l'enquête. C'est le cas des objets composant les réseaux *Philips* et *Sen.se*. Cependant, des données pertinentes sont contenues dans les passerelles. Elles sont liées à l'activité du réseau et aux configurations du système. Certains objets sont plus polyvalents dans leur fonctionnement. C'est notamment le cas des objets du premier réseau étudié et de l'écosystème *Orvibo*. Ainsi, chaque objet doit être traité individuellement. Dans le premier réseau, les données de l'*Apple Watch* et du pèse personne *Nokia* sont automatiquement synchronisées avec l'*iPhone*. Pour le *Terraillon Dot*, une action manuelle de synchronisation doit être effectuée sur l'application par l'utilisateur. Il est à noter quelques redondances d'information entre celles remontées dans le système d'information et les données stockées localement, tant qu'il n'y a pas eu de réécriture en mémoire. Par conséquent, il est pertinent en fonction des cas d'usage, de l'accessibilité de la donnée et de la proximité par rapport à l'incident, d'agir sur les objets directement. Cette opération est réalisée en situation dégradée et d'urgence. Le *smartphone* concentre beaucoup d'informations pertinentes en agissant comme une passerelle (zone 1) et une interface utilisateur dans la gestion des objets (zone 3). Pour le réseau *Orvibo*, tous les objets se synchronisent automatiquement. Cependant, ils ne disposent pas de la même politique de gestion des données. Les capteurs de mouvement et d'ouverture ne contiennent aucune donnée en mémoire utile ou pertinente pour l'enquête. La caméra *Orvibo* embarque un stockage externe sous la forme d'une carte *Secure Digital* (SD). Elle sert de mémoire tampon avant transmission de l'information et en cas de perte du réseau.

Les six autres appareils des réseaux 2, 3 et 4 sont traités indépendamment. Les objets *Heroz*, *Sens'it* et la caméra *IP M136W* se synchronisent automatiquement avec le réseau. Elles ne contiennent que peu de données pertinentes au regard de leurs capacités de stockage et leur fonctionnement. L'*Amazon Echo*, le *Raspberry Pi* et le *WinkHub* synchronisent certaines de leurs données avec le réseau. Ainsi, ils conservent les informations pertinentes localement, en particulier des données de l'environnement et des informations à caractère personnel renseignées par l'utilisateur. Elles doivent donc être traitées individuellement par les enquêteurs.

6.3.3 Analyse de la topologie du réseau, extraction des équipements et placement sous-scellé

Le réseau 1 est structuré selon une relation point à point. Chaque objet est indépendant. Ainsi, les objets connectés ont été isolés individuellement en désactivant la connexion Wi-Fi et Bluetooth sur le *smartphone*. L'extraction des données du pèse personne *Nokia* et du

Terraillon Dot ne peut être accomplie en dehors d'un cadre laboratoire. L'opération nécessite l'usage d'équipements spécifiques d'extraction. La balance et les analyseurs de sommeil sont éteints en retirant leur alimentation électrique interne, accessible pour un enquêteur. En-effet, les données présentes dans les mémoires volatiles de ces équipements ne sont pas pertinentes au regard des besoins de l'enquête. L'extraction des données de l'*Apple Watch* et de l'*iPhone* est réalisée avec des outils des sciences forensiques standards de type mallettes d'extraction et d'analyse. Ces appareils sont ensuite éteints pour être placés sous-scellés. Tous les objets du réseau 1 sont conditionnés de manière indépendante pour être analysés ultérieurement dans un espace adapté.

Le réseau 2 se compose d'un réseau principal et de réseaux périphériques. Le réseau principal est un réseau maillé hiérarchique. Les trois réseaux périphériques indépendants sont structurés selon une logique en étoile : *Orvibo*, *Philips* et *Sen.se*. Les objets *Orvibo* et *Philips* s'appuie sur le protocole ZigBee pour échanger avec leurs passerelles. L'écosystème *Sen.se* est basé sur un protocole propriétaire. Les passerelles jouent un rôle hybride en tant qu'interface entre les réseaux et les protocoles de la zone 1. Dans un premier temps, le lien entre la passerelle principale et chaque environnement est coupé. Ensuite, tous les objets sont déconnectés de leurs passerelles respectives. L'opération est réitérée pour la caméra *IP M136W*, l'*Amazon Echo* et le *Raspberry Pi*. L'extraction des données des objets connectés *Philips*, *Orvibo* et *Sens'it* et de l'*Amazon Echo* est opérée ultérieurement dans un espace adapté. Elle nécessite l'usage d'un équipement spécifique dans le dessoudage et la lecture des mémoires de stockage. Cependant, l'extraction des données présentes sur les différentes passerelles est techniquement réalisable sur le terrain par des techniciens en nouvelle technologie. En effet, le hub *Philips* dispose d'une API dédiée et l'*Orvibo* un Telnet ouvert. Néanmoins, l'accès nécessite de connaître les caractéristiques de connexion. Ces éléments sont renseignés dans la base technique de connaissance mise à disposition des unités techniques. La caméra *IP M136W* ne possède aucune donnée pertinente stockée localement. Les cartes de stockage externes de la caméra *Orvibo* et de le *Raspberry Pi* sont également récupérées. En l'absence d'information pertinente dans les mémoires volatiles respectives, les appareils sont éteints en retirant leur batterie interne ou leur branchement électrique au secteur. Tous les objets du réseau 2 sont scellés de manière indépendante pour être analysés ultérieurement en laboratoire.

Le réseau 3 est déjà isolé de tout réseau. Il est considéré comme un dispositif dit « hors ligne ». Le réseau 4 est structuré selon une topologie en étoile inversée. Localement, l'objet de ce réseau communique avec des passerelles privées externes. Il doit donc être isolé du réseau par l'usage d'une cage de Faraday en l'absence d'un moyen d'arrêt.

Certains appareils ne sont pas toujours collectés conformément à une stratégie d'investigation définie par les enquêteurs. Cependant, il est nécessaire de connaître leurs rôles et leurs identifiants, comme par exemple l'identifiant FCC ID ou le numéro de série (cf. chapitre 4).

Ces informations sont exploitées en fonction des besoins de l'enquête lors de la phase d'analyse ou au travers des réquisitions auprès des opérateurs de plates-formes IdO.

6.3.4 Collecte à l'épreuve de l'opérationnel

Plusieurs opérations sont effectuées successivement pour collecter les appareils numériques et leurs données : une rupture des liens de radiocommunication, le retrait des cartes externes, l'extraction des données sur le réseau ou les équipements, une modification de l'état des appareils et leur conditionnement sous la forme d'un scellé. Ces actions sont génératrices de traces pouvant aller d'une simple écriture à une altération des équipements, voire même une perte irréversible de données.

Les objets connectés dépendent d'une infrastructure locale. Ils forment un tout connecté. Une modification de l'infrastructure entraîne une écriture dans les journaux d'événement et dans le système. En rompant les liens physiques ou radio, l'écriture est localisée et la fuite des données est limitée. L'étude des passerelles *Sen'se*, *Philips* et *Orvibo*, n'a pas révélé d'écriture dans les journaux une fois les équipements déconnectés. Cependant, un événement est créé lors de la déconnection de l'*Amazon Echo* dans le répertoire : */system/dropbox/*. Cette entrée n'a aucun impact sur les données stockées initialement. L'opération est renseignée dans la traçabilité des opérations afin de justifier l'événement lors de la phase d'analyse et par la suite à une cour de justice.

L'extraction de données génère des traces et éventuellement des pertes d'information. L'acquisition directe laisse une écriture dans la mémoire vive. Elle modifie une source potentielle de données pertinentes, comme par exemple un élément de connexion. De plus, le résultat d'une acquisition en direct n'est ni répétable, ni reproductible. La seule méthode d'extraction de la mémoire vive sans altération est le Crash Dump. Malheureusement, cette approche ne peut être activée manuellement. L'acquisition de la mémoire interne pendant l'exécution entraîne également une altération plus ou moins importante du support. Certaines méthodes provoquent le démarrage ou la modification du système pour augmenter les privilèges ou exploiter une faille de sécurité. Afin de contrôler l'impact de ces solutions, les opérations sont effectuées dans des environnements contrôlés de laboratoire. Elles sont également jouées en amont sur des équipements ayant des caractéristiques identiques. Toutes ces opérations sont tracées dans la procédure par le technicien. Elles respectent et s'appuient sur des protocoles normés et discutés par la communauté de la criminalistique numérique, issus de retour d'expérience.

Un arrêt propre du système d'exploitation génère nécessairement une écriture dans le système. L'écriture est liée à l'enregistrement des données de l'application dans la mémoire vive. Inversement, un arrêt brutal consiste à couper l'alimentation électrique. Il protège la

mémoire des nouvelles entrées, mais entraîne une perte des données de la mémoire vive. La mise en œuvre d'un arrêt propre ou brut du système est susceptible d'entraîner une écriture dans les journaux d'événements, la suppression de fichiers temporaires et la purge des caches. Elle peut être accompagnée de l'exécution d'un script ou d'une application conduisant à l'effacement ou au chiffrement des données. Si la machine est dotée d'une interface homme-machine, il est nécessaire de capturer l'horodatage de la machine, le réseau auquel elle est connectée et les applications qui s'exécutent en arrière-plan. Ces actes techniques doivent être effectués avant d'éteindre et de sceller le dispositif connecté.

Dans certains cas, l'objet ne peut pas être désactivé ou arrêté. Ainsi, il continue à vivre et à écrire des données en mémoire, entrainant un risque de réécriture sur des éléments importants de l'enquête. C'est notamment le cas avec une montre GPS active, telle que l'*Apple Watch* ou une montre sportive de marque *Garmin*. Il est donc plus intéressant de limiter son fonctionnement. L'enquêteur est amené à arrêter des applications ou des fonctionnalités actives pour geler l'état de l'objet, comme par exemple une désactivation du GPS ou l'arrêt d'une application santé. À la situation du maintien allumé des objets scellés se pose la question de l'alimentation électrique des équipements. L'arrêt des applications limite également la consommation. Ces opérations techniques sont nécessairement tracées dans la procédure.

6.4 En quelques mots : collecte des traces numériques dans un environnement IdO

L'Internet des objets est composé de dispositifs hétérogènes et interconnectés du réseau. Une manipulation non raisonnée est source de dommages irréversibles tels que des altérations ou des destructions de données. Les caractéristiques techniques des équipements, de la topologie du réseau, des dépendances et de la politique de gestion de la donnée influencent et déterminent le processus de collecte des contenus et des contenants numériques et conditionnent leur placement sous scellé.

À l'issue de la phase de collecte, les enquêteurs procèdent à l'extraction et à l'analyse des données contenues dans les équipements. Ces actes techniques sont opérés directement sur la scène de crime ou dans un espace dédié en fonction des caractéristiques des matériels à étudier. L'analyse va chercher à établir une cohérence entre la donnée provenant des objets connectés et les hypothèses émises. La donnée devient une information contextualisée donnant la possibilité aux enquêteurs de reconstruire la chronologie des faits survenus.

Chapitre 7

Analyse forensique des traces numériques

L'Internet des objets contribue à l'apport subséquent de traces témoignant d'une activité ou de faits révolus. La pleine réponse criminalistique s'avère délicate face à la structuration étendue de cet espace numérique. Les traces sont dispersées que ce soit localement mais également que ce soit par des ramifications de l'infrastructure connectée et des espaces de traitement en ligne. La donnée est susceptible d'être partielle dans l'objet local mais devient un ensemble cohérent dans l'arborescence numérique. Cette détermination de la présence et du positionnement de l'information demeure unique à chaque écosystème. Elle est en particulier liée à la politique de gestion de la donnée, tant au niveau du stockage que de sa synchronisation avec le réseau. L'analyse des traces est donc beaucoup plus complexe que dans le cadre de la criminalistique numérique traditionnelle, en raison de son caractère multidimensionnel et pluridisciplinaire. Elle s'accompagne également d'un travail de contextualisation de la donnée et de compréhension de sa diffusion, selon les facteurs espace et temps.

7.1 Problématique de l'analyse dans l'Internet des objets

L'analyse des objets connectée et de ses artefacts n'est pas une science nouvelle. Pour apporter une réponse globale, elle ne doit pas se limiter uniquement aux équipements. L'objet connecté s'inscrit dans un plus vaste ensemble.

7.1.1 Artefacts de l'Internet des objets

La littérature scientifique est riche de nombreux travaux dans l'étude des objets connectés et de leurs artefacts, que ce soit à partir des montres ou des bracelets connectés [131, 132, 133, 134, 135, 136, 137] ou des assistants vocaux [138, 139, 140, 141, 142, 143] et plus globalement de tout dispositif connecté de la vie quotidienne [144, 145, 146]. Ces dispositifs se caractérisent par leurs propres formats de données, protocoles et interfaces physiques [114]. Certains articles se focalisent sur l'analyse des données générées lors de l'usage des applications mobiles, comme par exemple [147, 148]. Ils font notamment référence aux données synchronisées, aux bases de données SQLite et aux fichiers de cache contenant des informations de connexion aux plates-formes de l'IdO. Des travaux portent également sur l'appréhension des journaux d'événement [149, 150] et sur l'analyse du trafic réseau en local [151, 72, 152, 153], en particulier dans le cadre des détections d'intrusion [154, 155].

7.1.2 Limites d'une approche traditionnelle et unitaire

La plupart des études demeurent imparfaites en se cloisonnant à un objet spécifique ou bien à une architecture locale constituée d'équipements de la même famille telle qu'une maison connectée et de sa domotique [156]. Ces approches omettent les compatibilités de connexion inter-objets et les nouvelles dépendances entre les systèmes ainsi que la dispersion de l'information dans l'infrastructure au gré des configurations et des services proposés. Un appareil est susceptible d'être contrôlé ou accessible à partir d'un matériel distinct du système, selon une structure de « lien caché ». Ainsi, la donnée se propage et est stockée dans les équipements du réseau, concourant ou non à l'objet cible [126]. Or, l'Internet des objets mixe de plus en plus les équipements connectés de différentes familles à partir de modules polyvalents. Les écosystèmes se personnalisent en fonction des choix et des configurations de l'utilisateur. Par exemple, les assistants vocaux et leurs solutions de commandes vocales natives relient des objets connectés d'une même maison qu'ils soient de la domotique, de l'électro-ménager et de la sécurité. Initialement relevant de la catégorie de simples équipements connectés, ils sont devenus de véritables écosystèmes évolués doués d'une certaine intelligence. Ce service commun est déployé depuis peu dans des objets hors du périmètre de la maison par l'intermédiaire de bracelets connectés ou de systèmes complexes tels qu'un véhicule connecté. Dernièrement, après des essais concluant chez *Ford* et *Mercedes*, *Amazon* vient de lancer un *Software Development Kit (SDK) Alexa Auto* téléchargeable sur *GitHub* intégrant cette fonctionnalité. De la même manière que le multimédia d'un véhicule connecté contient les données d'un téléphone synchronisé, ce système embarqué fortement contraint est susceptible de détenir des informations de la maison ou d'un individu et *vice-versa*. La couche application constitue le liant dans l'échange de l'information utile entre les différents environnements matériels, mélangeant des

données de tous horizons. Les frontières entre les systèmes connectés deviennent de plus en plus poreuses. Le marché tend à évoluer en ce sens au vu du développement des protocoles de communication autour de l'interopérabilité et la création de partenariat entre les entreprises de l'IdO, notamment par les plates-formes en ligne ou des solutions mutualisées telles que les assistants personnels virtuels. Cette évolution concourt à la création d'écosystèmes connectés polymorphes, mouvant au gré des configurations des utilisateurs. Cette problématique soulève plusieurs questions dans le domaine de la criminalistique moderne concernant le partage et le recoupement des informations utiles afin de reconstituer avec fidélité la chronologie des événements dans son contexte. Elle met à mal une approche statique et unitaire de la scène de crime, en s'inscrivant dans une approche plus globale de la chose. La valeur ajoutée de l'Internet des objets vient du fait que le tout et plus grand que la somme des parties, ce qui explique que les approches unité par unité passe à côté de la valeur ajoutée de l'IdO.

7.2 Collecte des données dans l'Internet des objets

La collecte des données dans l'Internet des objets s'intéresse à des données présentes localement mais également à l'extérieur de la scène de crime. Elle fait appel à la criminalistique des réseaux de communication, de la mémoire, des données et des fichiers supprimés et/ou morcelés, des systèmes embarqués et mobiles, de l'informatique nuagique et de la rétro-ingénierie.

7.2.1 Extraction des données locales

Il existe deux approches dans l'extraction des données sur des équipements physiques : « *live forensic* » et « *post-mortem* ». Ce choix est conditionné par l'état du matériel, les connaissances et les moyens techniques ainsi que l'effet recherché par l'enquêteur.

L'extraction « *post-mortem* » des données présentes au sein des objets connectés et des passerelles s'appuie sur les connaissances et les techniques de laboratoire développées dans le cadre d'un équipement mobile et d'un système électronique embarqué. Il existe plusieurs niveaux d'extraction offrant l'accès à des informations différentes et complémentaires : manuelle, logique et physique [157, 158] (cf. Figure 7.1).

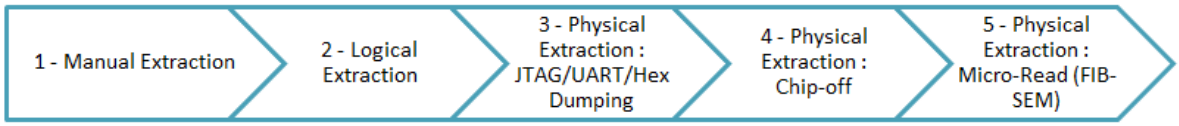


FIGURE 7.1 – Vision générale des niveaux d'extraction d'un objet connecté

7.2.1.1 *Live forensic*

Lors de l'appréhension de la scène de crime, l'enquêteur est à même d'opérer des acquisitions de données en temps réel sur des systèmes vivants, selon une démarche « *live forensic* ». Elle repose en la capture de la mémoire vive RAM ou des données issues des processus en fonctionnement associés à une session de communication [122]. Un certain nombre d'outils d'acquisition de la mémoire sont proposés dans la littérature. Ils s'appuient par exemple sur les vulnérabilités liées aux fonctionnalités de gestion de la mémoire du système comme avec le pilote du sous-système de mémoire partagée d'un Android (*Ashmem - Anonymous SHared MEMory system*) [159] ou bien de gestion des processus avec *Android low memory killer* [160]. Ils exploitent également les commandes d'accès et de chargement de la mémoire tels que les outils *fmem* ou *crash* du projet *Redhat* sur des noyaux Linux [161]. Cependant, cette acquisition de la mémoire en direct demeure complexe et limitée. Par exemple, des technologies de protection telles que le *Memory Protect Unit* (MPU) conditionnent les accès à des instructions ou des codes spécifiques. En outre, de nombreux équipements embarquent des techniques anti-forensiques susceptibles de déclencher des opérations d'écrasement de données et de métadonnées. Par exemple, *TimeStomp2* s'utilise dans la réécriture des horodatages dans un système de fichier *New Technology File System* (NTFS). Malgré un certain nombre d'outils développés pour l'acquisition de la mémoire vive à partir d'ordinateurs et des téléphones portables, tels que *dd*, *dumpit.exe*, *memoryze*, *nigilant32*, *readline*, *winhex*, *winen*, il n'existe que peu de solutions à ce jour adaptés aux dispositifs connectés de l'Internet des objets.

7.2.1.2 Extraction post-mortem manuelle

L'**extraction manuelle** des données est réalisée sur les équipements disposant d'une IHM. Elle consiste à faire des copies lors d'une manipulation manuelle de l'interface utilisateur par le biais de la photographie ou de la vidéographie. Des outils criminalistiques industrialisent cette fonctionnalité tels que *UFED Camera*, *XRY Camera*, *ZRT 3*, *Eclipse 3 Pro Kit*, etc. Cette technique d'extraction est chronophage pour les enquêteurs et n'autorise pas la récupération de données supprimées ou inaccessibles par l'utilisateur telles que des données système. De-même, elle n'est pas applicable lorsque l'accès est verrouillé et inconnu ou lorsque l'écran est endommagé. Elle génère également une écriture lors de la manipulation des fichiers. Toutefois, l'extraction manuelle est souvent utilisée sur le terrain pour orienter des investigations, pour figer un état du système lors de la collecte ou pour vérifier et valider d'autres niveaux d'extraction automatique.

7.2.1.3 Extraction post-mortem logique

L'**extraction logique** [162] cherche à trouver les éléments visibles à partir du système de fichiers. Cette opération nécessite de connaître au préalable les caractéristiques techniques et la version du système d'exploitation (OS) du dispositif. Ces informations conditionnent directement le choix de la stratégie de communication à adopter pour procéder à l'extraction logique : une connexion au bus universel en série *Universal Serial Bus* (USB), l'utilisation des protocoles séries ou sans-fil, une interface de programmation d'applications (API), des commandes propriétaires, etc. Ce moyen d'accès s'appuie sur les protocoles logiques de communication entre le dispositif cible et l'environnement d'analyse. Ainsi, en utilisant l'API, l'enquêteur échange directement avec le système d'exploitation de l'appareil. Cependant, il n'est capable que de requêter une extraction de données accessibles uniquement par le système d'exploitation. Dans certains cas l'équipement est placé en « mode diagnostic ». Cette opération offre un accès direct au système. C'est notamment le cas de certaines versions de l'*Amazon Echo* ou de l'*Amazon Fire TV* en activant l'*Android Debug Bridge* (ADB). Les données sont alors récupérées en utilisant le protocole du fabricant.

7.2.1.4 Extraction post-mortem physique

L'**extraction physique** est une copie bit-à-bit d'un support. Elle se décline en trois niveaux d'abstraction : le *Joint Test Action Group* (JTAG) ou *Universal Asynchronous Receiver Transmitter* (UART), le dessoudage suivi de la lecture du composant mémoire et la micro-lecture *Focused Ion Beam - Scanning Electron Microscope* (FIB-SEM). Cette opération d'acquisition réside en une copie de très bas niveau de toutes les données binaires physiquement présentes dans le silicium de la mémoire de l'équipement. Elle se traduit par une lecture électronique de l'état de toutes les cellules élémentaires de mémoire. Ainsi, elle offre une collecte des informations encore physiquement présentes dans la mémoire flash. C'est la différence cruciale avec l'extraction logique qui se concentre sur l'espace alloué. Il est intéressant de réaliser cette opération sur des mémoires telles que les cartes SD, afin d'étudier les espaces non allouées susceptibles de contenir de la donnée laissée à la suite d'un repartionnage de la mémoire.

7.2.1.5 Extraction post-mortem appliquée au cas d'étude

Dans notre cas d'étude, des extractions logiques et physiques des données présentes dans les équipements sont opérées (cf. Tableau 7.1). Le choix du type d'extraction est motivé par les informations que l'enquêteur souhaite récupérer, par la faisabilité technique de l'opération, par la durée et le coût de l'intervention. Plus le niveau d'extraction est élevé, plus le coût des outils et du processus d'extraction est important. À ce constat s'ajoute l'usage de tech-

niques invasives et destructives nécessitant un haut niveau expertise. En fonction des besoins d'enquête, il est parfois pertinent de réaliser plusieurs niveaux d'extraction pour obtenir des données brutes de l'ensemble de la mémoire et celles accessibles et interprétées par le système. En-effet, les méthodologies de l'analyse incluent la recherche par mot-clef sur les partitions allouées par le système et sur les espaces non alloués. Elles s'intéressent également à la récupération de fichiers supprimés et à l'extraction des informations de registre telles que les comptes ou les périphériques associés.

No	**Équipement**	**Fabricant**	**Type d'acquisition**
4	Caméra (Carte SD)	Orvibo	Extraction physique
5	Passerelle	Orvibo	Extraction physique (UART + JTAG 8P et Chip-off) Extraction logique (Telnet)
8	Passerelle	Philips Hue	Extraction physique (JTAG 14P et Chip-off) Extraction logique (API)
13	Mother	Sen.se	Extraction physique (Chip-off) Extraction logique (API)
14	Echo Spot	Amazon	Extraction physique (UART, JTAG 5P et Chip-off) Extraction logique (ADB)
15	Pi0 (carte SD)	Raspberry	Extraction physique
17	WinkHub2	Wink	Extraction physique (JTAG, UART et Chip-off) Extraction logique (API)
18	Watch 3	Apple	Extraction physique (IBUS S2) Extraction logique (Backup)
19	iPhone SE	Apple	Extraction physique Extraction logique (Backup)
20	Dot	Terraillon	Extraction physique (JTAG 6P et Chip-off)

TABLE 7.1 – Type d'acquisition opérée sur les dispositifs de la scène de crime

Les mémoires des objets connectés *Sen.se*, *Orvibo*, *Philips*, *Sens'it*, du bracelet *Heroz*, de l'*IP Camera M136W* et de la balance *Nokia* n'ont pas été traitées en raison d'espaces de stockage interne de faible dimension, d'une remontée automatique de la donnée au sein du réseau et de leur pertinence au regard de l'enquête tant dans la proximité par rapport à l'événement recherché que dans leur position sur la scène de crime. Leurs éléments d'identification sont néanmoins exploités dans le cadre de réquisitions auprès des opérateurs de plates-formes et dans la caractérisation des événements lors de la phase d'analyse.

L'automatisation du décodage et le formatage des résultats de l'extraction sont facilités par l'emploi des outils classiques de la criminalistique numérique tels que *X-Ways*, *Forensic Explorer*, *OSForensics*, *UFED*, *XRY*,*EnCase*, *FTK*, *Autopsie*, etc. Toutefois, leurs usages nécessitent des données ou des images dans un format standard ou reconnu. Ces solutions logicielles offrent également la possibilité d'effectuer des recherches Unicode ou *American Standard Code for Information Interchange* (ASCII). L'absence d'une structure de fichier standard rend le processus d'examen plus difficile et chronophage. Elle nécessite un profilage des données et des fichiers. Elle est souvent accompagnée par un travail de rétro-ingénierie, en particulier face aux mécanismes de compression ou de chiffrement des données.

7.2.2 Extraction des données externes

La communauté scientifique définit l'informatique légale dans les nuages « *Cloud Computing Forensic* » comme l'application de la criminalistique numérique dans des environnements en nuage [75, 163, 164]. Techniquement, il s'agit d'une approche médico-légale hybride dans la recherche et la découverte de preuves numériques. Elle relève d'environnements distants et virtuels, mais directement accessibles au travers d'un réseau, de clients légers et lourds. Sur le plan organisationnel et juridique, elle implique des interactions entre les services d'enquête et les opérateurs de plates-formes gestionnaires de l'infrastructure, mêlant des situations multi-juridictionnelles et multi-locataires. Le NIST [165] par son groupe de travail *Cloud Computing Forensic Science Working Group* (FSWG) a publié en août 2020 des recommandations en la matière.

La collecte de données dans le cadre de l'informatique légale dans les nuages se heurte à des défis, que ce soit en la localisation des artefacts, des données volatiles, mais également que ce soit en la collecte de données visibles ou supprimées sur des machines virtuelles partagées et distribuées. À ces difficultés s'imbriquent la question de l'intégrité des informations issues d'un environnement à locataires multiples où les données sont partagées entre plusieurs serveurs situés dans plusieurs endroits et accessibles par plusieurs parties. L'analyse des données extraites va s'atteler à corréler et à imbriquer des artefacts provenant de différentes sources en reconstituant les événements à partir d'images virtuelles ou de stockage. Elle s'intéresse également à leurs synchronisations et à définir une chronologie des données des journaux d'événements. Dans le cas d'étude, plusieurs plates-formes propriétaires centralisent la remontée de l'information : *Amazon*, *Philips*, *Orvibo*, *Sen.se*, *Apple*, *Terraillon*, *Withings*, *Heroz*, *Sens'it* et *Wink*. En fonction de la configuration définie par l'utilisateur, un dispositif connecté localement fait appel à des services concourants en ligne. Par exemple, les données *Sens'it* sont susceptibles d'être retrouvées sur un *Google Cloud* ou les données *Philips* sont possiblement

partagées avec le *Cloud Amazon*. La collecte de l'information utile est donc conditionnée à une connaissance fine des équipements locaux et de son infrastructure.

7.3 Exploitation et analyse des données recueillies

Les traces obtenues résident en de nombreux fragments contenus dans une pluralité de supports d'un même réseau, susceptibles d'évoluer dans le temps et l'espace. Afin qu'elle ne soit pas parcellaire, l'analyse ne doit pas se focaliser sur l'étude d'un unique objet mais sur l'écosystème dans son intégralité. Elle consiste donc à étudier et à penser la donnée selon trois axes : le **temps** en définissant la chronologie des événements et des phénomènes itératifs, l'**espace** en positionnant la donnée dans l'infrastructure et en tenant compte de l'environnement local et le **contexte** en analysant l'événement eu égard des rôles et des actions des différents équipements face au phénomène. Par cette approche ternaire, l'enquêteur cherche à déterminer le cycle de vie de la donnée afin de la qualifier et d'établir sa cohérence. Nous partons du postulat qu'il y a une dépendance de causes (cf. chapitre 2 - principe d'abduction) entre les différents événements et l'état de la donnée dans le système [166].

Dans le but de procéder à une analyse pertinente des événements, il est nécessaire d'observer l'information selon un horodatage commun. L'enquêteur doit veiller à récupérer l'horodatage de chaque équipement ou la méthode de synchronisation des systèmes. Il calcule l'écart entre cette donnée matérielle et l'horloge universelle afin d'intégrer dans son raisonnement ce différentiel.

7.3.1 Type de données retrouvées dans les équipements connectés

La pluralité de l'écosystème connecté offre une variété de données produites et échangées. Cette source d'information se décline en « données fonctionnelles » utiles au rendu du service (contenu multimédia, données de téléphonie et de localisation, historique web, mesures d'environnement et d'activité) et en « données périphériques » liées au fonctionnement du système et du réseau (journaux d'événements). Les équipements locaux comportent également des données de contexte telles qu'une configuration physique et logique d'un lieu, une habitude de vie, un enregistrement sonore ou vidéo d'un phénomène ponctuel. Ces éléments sont combinées à des données à caractère personnel telles que l'identité du consommateur de service, son profil numérique et biométrique. Ces informations offrent aux enquêteurs de qualifier un phénomène et de restituer avec fidélité la succession des événements dans son environnement.

Les passerelles domotiques *Orvibo* et *Philips* contiennent des informations sur l'utilisateur, sur le réseau ZigBee et son adressage, sur la configuration physique et logique de l'habitation

avec ses étages et ses pièces, sur des familles d'objets connectés et leurs associations avec l'environnement, sur les usages programmés et les scénarios de sécurité. Elles enregistrent l'ensemble des interactions avec l'infrastructure, catégorisées selon le type de déclenchement d'une action : une commande générée par une application mobile ou par des objets locaux, mais également des scénarios programmés en réponse à un phénomène constaté. Toutes ces transcriptions sont horodatées et renseignés. D'un point de vue criminalistique, ces informations apportent à l'enquêteur des renseignements sur le « quand » et le « comment » une présence ou une activité détectée par l'écosystème et sur l'usage domestique des objets de vie. Il obtient une première identification des acteurs du phénomène ponctuel et caractérise son déclencheur. L'application mobile enrichit l'investigation par des données inscrites dans le temps en particulier pour comprendre le contexte mais également sur la situation du phénomène avec les dates des interactions et l'association matérielle. La carte de stockage externe de la caméra *Orvibo* complète les rapprochements par un retour multimédia.

Les capteurs *Cookie* mesurent des changements d'état comme un mouvement ou une température. La passerelle *Mother* contient les journaux d'événements retraçant les communications et la remontée de l'information vers l'infrastructure connectée. L'application *Sen.se* associe une mesure de phénomènes, avec une information prédéterminée et configurée. Elle rassemble des données sur l'utilisateur, sur le réseau propriétaire et sur la logique d'association entre un objet et une action via des rendus graphiques. Les enquêteurs obtiennent de cet écosystème une chronologie de phénomènes survenus en cohérence avec l'usage d'un objet physique, digitalisé par le *Cookie*.

Dans notre cas d'usage, l'*Amazon Echo* n'est pas couplé à une gestion d'un parc d'équipements connectés en direct mais à un rôle d'interface intelligente délivrant des commandes à l'infrastructure. L'utilisateur est autorisé à activer les ampoules connectées par une commande vocale. Cette interface contient principalement des informations sur l'utilisateur, le réseau et un ensemble de logs d'activités liées au système, au fonctionnement courant, aux événements et aux échanges avec le réseau. L'application mobile concentre l'historique des interactions et les requêtes de l'assistant vocal. Ainsi, les journaux d'événements avisent l'action de capture d'un son au moment des faits. Cependant, les enregistrements sonores sont stockés sur la plate-forme *Amazon*. Seuls leurs liens d'accès sont renseignés dans l'application.

L'analyse du *WinkHub* est pertinente pour comprendre l'architecture du réseau local. Cette passerelle contient l'ensemble des événements réseau et les identifiants des équipements connectés. Elle identifie et cartographie le cheminement et la remontée de la donnée au travers de l'infrastructure connectée vers Internet. L'analyse du lien de dépendance avec l'*iPhone* est appropriée pour les investigations numériques. Cette connexion renseigne une proximité géographique du *smartphone*.

Le *Terraillon Dot* enregistre les mouvements sur une surface plane, ici le lit. Il contient l'historique des mesures réalisées dans un temps proche, les informations de configuration du réseau et de sa synchronisation. La remontée des données est déclenchée manuellement à partir de l'application mobile. Cette applicatif regroupe des données sur l'utilisateur comme son profil numérique et santé mais également des informations de réseau et de synchronisation. Les mesures sont disponibles sur une période de 30 jours : la durée de sommeil, l'événement de levé/couché, les mouvements du corps, la respiration, le rythme cardiaque, etc. L'*Apple Watch* enrichie et solidifie ces mesures en croisant des mesures d'activité physique de l'utilisateur couplée à des informations de localisation. L'enquêteur récupère une image fidèle des habitudes du bénéficiaire du service.

Les différents écosystèmes connectés de ce cas d'étude apportent une richesse d'information inédite, phénomène de numérisation d'un profil utilisateur dans le temps et l'espace en relation avec un contexte précis. La redondance et le croisement des données dans les différentes parties de l'infrastructure délivrent une certaine fidélité des mesures et une pondération des éléments configurés à la main de l'utilisateur. Ces éléments sont également à placer en perspective avec les données récupérées sur les plates-formes Cloud. Elles apportent le renseignement manquant dans la compréhension du phénomène révolu. Ces espaces véhiculent des associations inédites pour l'investigation judiciaire sur le profil numérique de l'utilisateur et son schéma relationnel. À titre d'illustration, la plate-forme *Amazon* contient les données relatives à la maison connectée mais également des liens avec d'autres objets rattachés au compte utilisateur, des habitudes de navigation, d'achat, des usages, des répertoires personnels, de la géolocalisation, etc. La seule contrainte dans le traitement de cette source d'information réside en des limites légales.

7.3.2 Traitement des liens cachés et des dépendances

Dans un écosystème connecté et synchronisé, un événement relève de plusieurs interactions entre des dispositifs numériques. L'information générée est diffusée à un réseau d'équipements interdépendants. Ainsi, cette donnée se retrouve potentiellement stockée sur un ou plusieurs supports situés à l'extérieur de l'écosystème primaire. En travaillant de façon unitaire sur chaque support, la corrélation entre les équipements est perdue. Afin d'appréhender cette question, l'enquêteur doit intégrer dans sa phase d'analyse l'étude de l'architecture réseau (*IoT network monitoring*) en identifiant les dépendances et les rôles des équipements dans la manœuvre. Il doit également modéliser l'activité en se basant sur l'étude des journaux d'événements des systèmes et des interactions.

7.3.2.1 Identification des équipements connectés

Dans les chapitres précédents, nous avons profilé les appareils physiques composant l'Internet des objets (cf. Tableau 7.2). De manière générique, l'environnement local est composé d'objets connectés et de passerelles. Les objets connectés sont plus ou moins élaborés. Certains se comportent comme de simples **capteurs** ou **actionneurs** en mesurant, détectant et réagissant à certaines données ou commandes de l'environnement physique. D'autres **objets** comportent plus de capacité de calcul et de stockage avec plus ou moins d'autonomie pour interagir avec le réseau, comme par exemple des caméras IP ou des téléviseurs intelligents. À ces caractéristiques fonctionnelles, s'ajoute un facteur de dépendance des objets pour communiquer. Ainsi dans le chapitre consacré à la collecte des contenus et des contenants (cf. chapitre 6), nous avons distingué les objets par rapport à leurs capacités à échanger de la donnée directement hors de leurs écosystèmes. Nous avons classé les passerelles en fonction de leurs rôles dans l'infrastructure : soit en tant que **nœud de réseau** ou soit en tant qu'**interface** avec le monde extérieur. Néanmoins, ces distinctions sont de plus en plus complexes et mises à mal par le développement de solutions hybrides, en particulier avec la décentralisation du traitement par les phénomènes de « *edge computing* » pour les passerelles et de « *fog computing* » pour les objets connectés. Par ailleurs, l'écosystème local est régi par un **contrôleur**. Il est potentiellement une partie intégrante d'une passerelle, comme dans le cas de stations connectés ou demeure un dispositif distinct, comme une interface applicative mobile. L'Internet des objets est complété par des **services Cloud** véritable carrefour de l'information.

Objet connecté	Capteur ou actionneur	Capteurs Sen.se (cookie), ampoules Philips, capteurs d'ouverture et de présence Orvibo caméra Orvibo, IP Camera M136W, Terraillon Dot, Sens'it, bracelet Heroz et balance Nokia
	Objet évolué	Amazon Echo, Raspberry Pi0 et Apple Watch
Passerelle	Nœud	Passerelles Mother, Orvibo et Philips
	Interface vers l'extérieur	WinkHub2 et iPhone
IHM	Contrôleur	Amazon Echo et applicatifs (iPhone)

TABLE 7.2 – Classement générique des équipements de la scène de crime

7.3.2.2 Étude des événements réseau et des liens cachés

L'activité réseau entre les objets, les nœuds et les passerelles s'accompagne de trois grandes familles de logs : la mise à jour des statuts des équipements, les commandes d'action et les informations relatives au fonctionnement du réseau (cf. Figure 7.2). La mise à jour des

statuts des équipements contient l'état des capteurs à un instant donné. Cette information est envoyée par les objets aux passerelles de façon séquentielle et à intervalle régulier. Le journal des commandes d'actions contient les commandes envoyées par les utilisateurs aux capteurs en direct ou à une passerelle. Les trames sont constituées au minimum d'un émetteur, d'un destinataire et d'une description de la commande. En analysant l'attribut de charge utile (format *JavaScript Object Notation* (JSON) ou *Extensible Markup Language* (XML)) et les valeurs intégrées des journaux, la commande de l'action vers les dispositifs est identifiée et extraite [167]. Le journal associé à l'activité réseau contient des informations relatives à l'état du réseau. Il trace les appairages et tous les échanges entre les équipements d'un même réseau. Il se résume sous la forme de trames contenant l'identification du dispositif et une valeur d'activité réseau. Les passerelles forment une communication bidirectionnelle qui agit comme un messager centralisateur faisant le lien entre les différents médias et les plates-formes Cloud. En enregistrant les activités du réseau qui passent dans les deux sens, les événements sont identifiables et extraits.

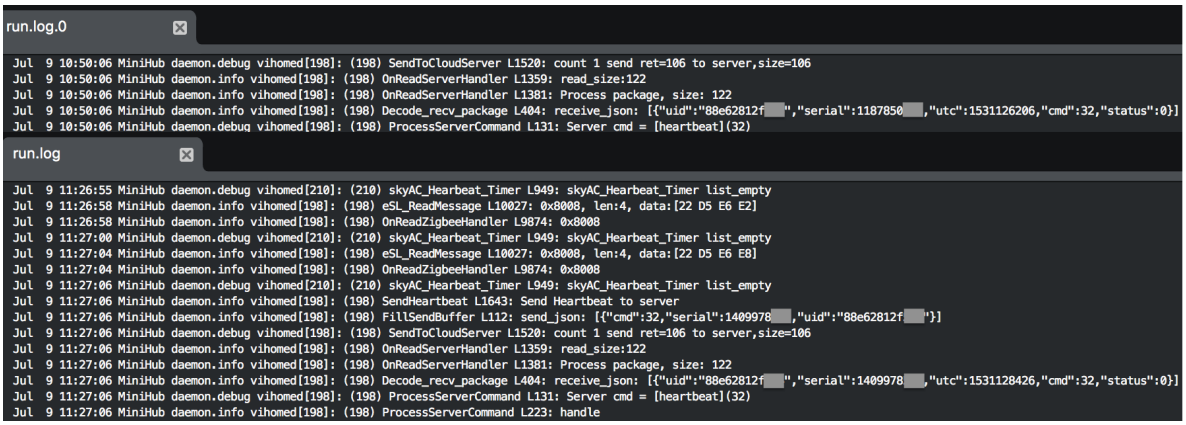


FIGURE 7.2 – Logs de la passerelle *Orvibo*

L'étude des liens cachés entre les équipements s'appuie sur la caractérisation des événements multiples partageant un attribut unique. Par exemple, en étudiant les journaux d'événements contenus dans la passerelle et l'application *Philips*, l'événement « lampe allumée » est horodaté. Cependant, il doit être caractérisé plus précisément. S'agit-il d'une action de l'utilisateur par le biais de l'application téléphonique, d'un interrupteur externe, d'un signal d'un capteur ou d'une commande vocale transmise par l'*Amazon Echo* ? A-t-elle été effectuée par un utilisateur connu ? S'agit-il d'une action programmée ? Pour chaque question, une donnée unique est associée. Elle caractérise l'événement qu'il soit dû à une interaction humaine active ou passive, sans interférence, direct ou non. Cette approche analytique est généralisable à tous les objets de l'infrastructure connectée. Elle permet de déterminer les liens cachés, ici une association entre l'*Amazon Echo* et la solution *Philips* à la suite d'une commande vocale

Alexa. Pour ce faire, l'événement est représenté à l'aide d'un modèle graphique (cf. Figure 7.3). Cette démarche simplifie le processus de corrélation et le regroupement des événements. En combinant l'horodatage, les groupes de même dimension temporelle sont rassemblés puis comparés en utilisant les attributs communs. De manière plus globale, nous obtenons une nouvelle modélisation du trafic réseau de la scène de crime basée sur l'orientation du réseau et de ses attributs (cf. Figure 7.4). Cette représentation graphique modélise le flux des messages de communication et aide à identifier les sources des preuves (cf. Tableau 7.3). L'enquêteur détermine les acteurs de l'événement, leurs positions, les actions effectuées ou détectées et la réponse des objets aux différentes sollicitations. Cet attribut unique se décline selon les identifiants d'un utilisateur, d'un objet, d'un lieu ou d'une donnée échangée. L'enquêteur va chercher à relier les différents équipements connectés par rapport à ce paramètre commun selon une approche heuristique. Cette démarche cherche à discriminer les éléments non pertinents et à orienter l'analyse.

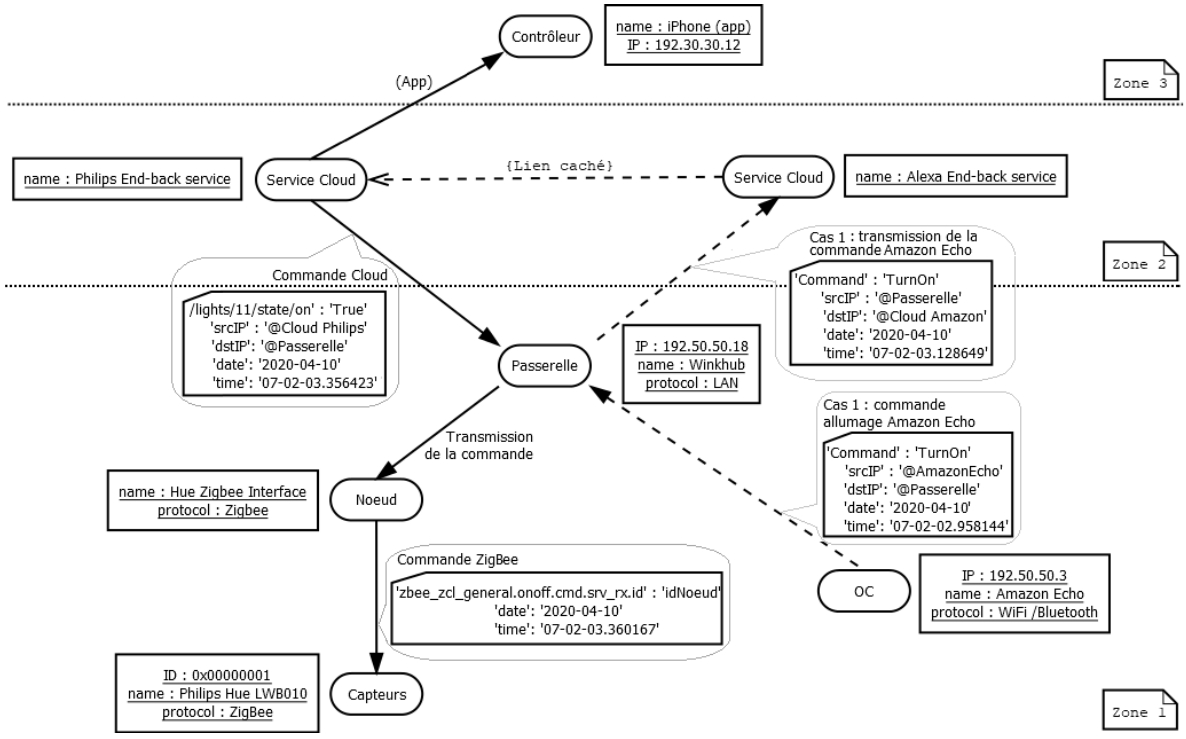


FIGURE 7.3 – Modélisation de l'action « allumage » d'une ampoule connectée de la marque *Philips* à la suite d'une commande vocale *Alexa*

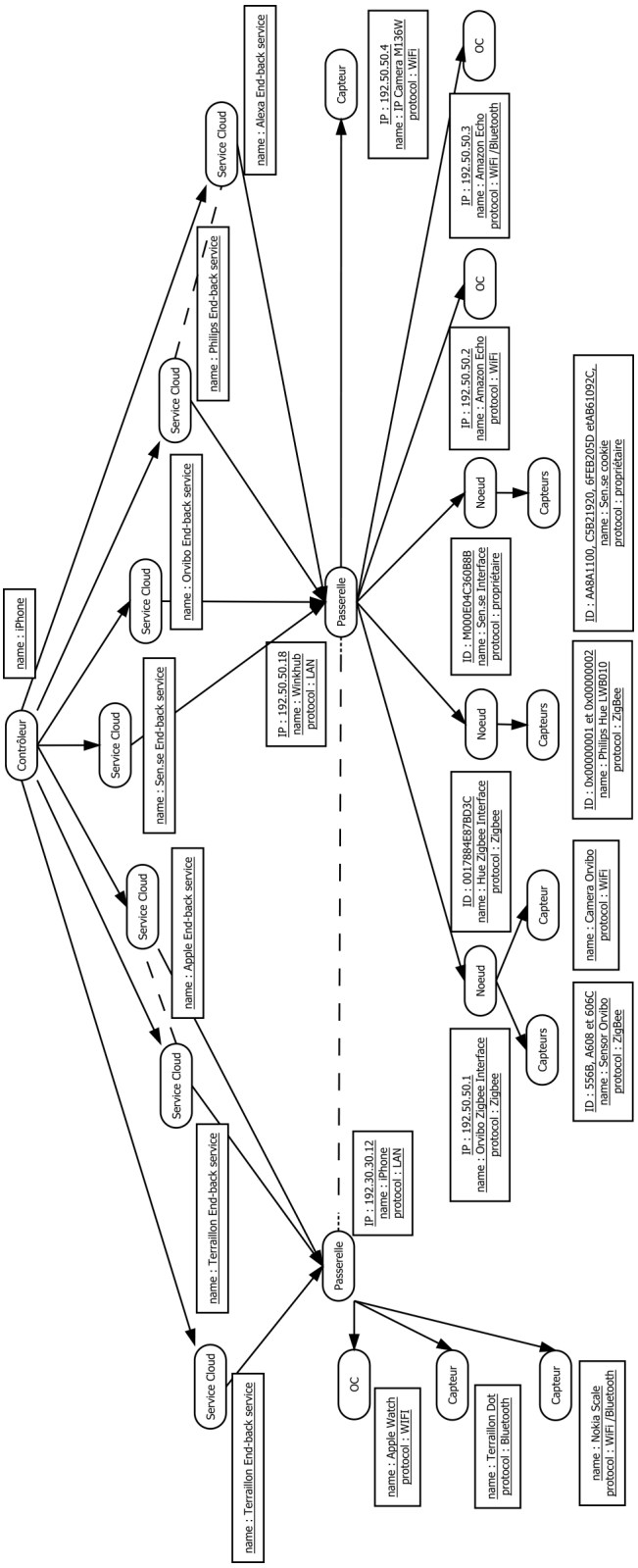


FIGURE 7.4 – Cartographie générale de l'environnement connecté

Source	Éléments	Chemin
Passerelle Philips	- Statut de connexion au Cloud - Date de la première et la dernière interaction réalisée par l'utilisateur - Liste des personnes autorisées à interagir	`/config/`
	- Evènements programmés ou configurés	`/schedules/` et `/scenes/`
Amazon Echo	- Informations sur le compte utilisateur (nom et type de service)	`/system/users/` : `0.xml (P16)` et `/0/accounts.db (P16)`
	- Logs d'activité	`/system/dropbox/`
Application Philips	- Enregistrement des actions effectuées par l'utilisateur à partir de l'application	`com.philips.lighting.hue2` : `/Library/com.amplitude.database/`
	- Logs d'activité	`group.com.philips.hue2` : `/debuglog/`
Application Amazon Echo	- Historique des discussions	`AlexaMobileiOSComms.sqlite`
	- Journal des cartes et historique des interactions vocales	`RCTAsyncLocalStorage_V1`

TABLE 7.3 – Éléments utilisés dans la reconstruction de la chronologie de l'événement « Lampe allumée »

À partir de notre cas d'usage, nous constatons que les activités du réseau présentent trois caractéristiques principales. Chaque dispositif a un ensemble de numéros de port fixes. Les équipements lancent des requêtes *Domain Name System* (DNS) pour un nombre limité de domaines correspondant pour la plupart aux noms de domaine de leurs fournisseurs. Ils utilisent un nom unique sous la forme d'un *User-Agent* et d'une adresse MAC.

7.3.3 Chronologie du phénomène criminel

À partir des données mises en perspective, l'enquêteur est en mesure de dater l'intrusion dans le domicile de la victime avec le capteur de la porte d'entrée. Il matérialise le parcours du mis en cause avec le système domotique *Orvibo*, corroboré par le mouvement du capteur *Sen.se Cookie* n°9. Il détermine l'heure du décès de la victime à partir des données de santé de la montre connectée (cf. Tableau 7.4). Il constate l'absence de modification de la scène de crime en particulier dans le déplacement post-mortem du corps en croisant les données du *Terraillon Dot* et de l'*Apple Watch*. Avec tous ces éléments, il est en mesure d'émettre des hypothèses sur le lieu du meurtre et les circonstances. Ces informations doivent être recoupées avec les données médico-légales recueillies sur la scène de crime et sur la victime.

Date	**Description**	**Source locale**	**Faits**
T0 10/04/20	Lampes éteintes (6 et 7) / Porte et fenêtre fermées (1 et 2) / Etat repos : activité cardiaque de l'Apple Watch (18) et Terraillon Dot (20)	Hub Philips et Orvibo, iPhone (App. santé et domotique) et Terraillon Dot	Présence d'une personne en pièce 2 (connue) et aucun mouvement détecté.
06:43:17	Ouverture de porte détectée (1)	Hub Orvibo, WinkHub et iPhone (App. domotique)	Présence d'une personne en pièce 2 (connue) et d'une personne en pièce 1 (non reconnue). Mouvements en pièce 1.
06:44:03	Détection du mouvement (3)		
06:44:12	Lancement de la Caméra Orvibo (4)	Caméra Orvibo (carte SD), Hub Orvibo, WinkHub et iPhone (App. domotique)	
06:52:46	Mouvements Cookie (9)	Mother, WinkHub et iPhone (App. Sen.se)	Présence d'une personne en pièce 2 (connue) et d'une personne en pièce 3 (non reconnue). Mouvements en pièce 3.
06:54:16	Déclenchement IP Camera M136W	WinkHub, iPhone (App. IP Camera)	
06:57:11	Mouvement Terraillon Dot (20)	Terraillon Dot	
07:01:04	Mesure d'une accélération du rythme cardiaque Apple Watch (18)	iPhone (App. santé)	
07:02:02	Commande vocale d'allumage (14)	Amazon Echo et WinkHub	
07:02:03	Allumage de l'ampoule Philips (6)	WinkHub, Hub Philips et iPhone (App. domotique)	
07:07:01	Mesure de l'arrêt cardiaque (8)	iPhone (App. santé)	Présence d'une personne en pièce 2 (connue).
07:07:54	Fin de la mesure du mouvement Terraillon Dot (20)	Terraillon Dot	
07:11:44	Détection du mouvement (3)	Hub Orvibo, WinkHub et iPhone (App. domotique)	Présence d'une personne en pièce 2 (connue) et d'une personne en pièce 3 puis 1 (non reconnue). Mouvements en pièce 3 et 1.
08:17:21	Détection du mouvement (3) : arrivée de la patrouille		
08:24:56	Détection du mouvement (3)		
09:12:00	Détection du mouvement (3) : arrivée enquêteur en nouvelles technologies		
T1	Lampes Philips : allumée (6) et éteinte (7) / Porte ouverte (1) et fenêtre fermée (2)	Hub Philips et Orvibo, WinkHub et iPhone (App. santé et domotique)	

TABLE 7.4 – Chronologie des événements d'une scène de crime

Par ailleurs, les données numériques sont en mesure d'orienter certaines investigations comme des relevés de traces papillaires et biologiques sur la scène de crime en cohérence avec le parcours criminel du mis en cause. Sur une habitation de plus grande taille, ces informations aident à délimiter et à discriminer une zone d'étude en définissant une stratégie d'investigation. Ainsi, les objets connectés donnent la possibilité de vérifier des hypothèses de travail en apportant de nouveaux éléments matériels, comme par exemple un mobile incohérent avec un mode opératoire. L'étude de la chronologie des événements peut également renseigner sur une éventuelle préméditation dans la logique criminelle.

7.4 En quelques mots : analyse de traces dans l'IdO

L'Internet des objets est une convergence de l'Internet et des réseaux de capteurs avec une vision de la communication de machine à machine. Cette structuration de l'écosystème connecté fournit une infrastructure commune tout en apportant une certaine cohérence de solutions par l'intermédiaire d'applications logicielles transverses telles que les « assistants personnels intelligents ». Cet ordonnancement offre à des entités du monde réel de créer, de partager et de valoriser la donnée numérique pour délivrer de nouveaux services.

Les équipements connectés sont des réceptacles d'information sans précédent, enrichie par la convergence des solutions. Ce phénomène génère des défis et des opportunités considérables pour l'investigation numérique en particulier concernant des dépendances de solutions et de liens cachés entre des objets hétérogènes. Afin de répondre avec pertinence aux besoins d'enquête et de comprendre les phénomènes révolus, le technicien en nouvelles technologies doit regarder la donnée provenant de multi-sources selon le prisme du temps, de l'espace et du contexte. Il s'appuie pour cela sur l'étude de l'architecture réseau, du rôle de chaque équipement en particulier dans les actions opérées et sur la compréhension de la migration des données au sein de l'infrastructure. Le croisement de traces obtenues intelligemment garantit des contrôles et des enquêtes inédites pour identifier les personnes, les lieux, les événements et les éléments associés à l'affaire judiciaire.

Les données parcellaires dans une multitude de dispositifs interconnectés s'inscrivent dans un tout lors de l'analyse. L'approche globale de l'écosystème connecté permet la capture de plus d'information que la simple somme des parties notamment à travers les mécanismes réseaux mis en œuvre.

Chapitre 8

Identification et caractérisation d'un objet connecté (principe de Kirk)

Dès lors que nous commençons à imaginer quelques scénarios d'investigation pour l'Internet des objets, nous constatons, compte-tenu de la diversité des objets et des protocoles utilisés, qu'il n'existe aucune approche unique et universelle permettant d'accéder aux données. Il est donc impératif de connaître la nature et le fonctionnement d'un objet afin de savoir dans quelle mesure il est possible d'en récolter de l'information, et, le cas échéant, de décider de la stratégie à adopter pour appréhender et accéder à ces données. Cette nécessité fait apparaître une problématique évidente : il n'est pas possible de connaître à l'avance le fonctionnement de tous les objets disponibles sur le marché. Cette affirmation s'explique par différentes contraintes : la variété des objets en constante augmentation diffus dans des domaines variés d'application, l'absence de renseignement sur des protocoles, les développements propriétaires et des usages détournés de solutions ou de fonctionnalités innées. Nous pouvons citer le cas d'une station d'optimisation de la consommation énergétique embarquant un capteur acoustique permettant de caractériser un niveau sonore et une présence. À ces difficultés, s'ajoutent les dimensions temporelles, financières et humaines de l'enquête, dans l'analyse d'une multiplicité d'objets. Pour pallier ce problème, il est nécessaire de mettre au point des techniques d'identification fine des équipements connectés basées sur des caractéristiques connues et communes à ces objets, et/ou dérivables d'un objet à un autre. Ces outils doivent être robustes, fiables et économiques, tout en évitant d'altérer les données probantes.

8.1 Techniques d'identification fine d'un objet connecté

La littérature scientifique est riche de travaux sur l'identification logicielle et matérielle des équipements électroniques. Elle se réfère aux différents attributs issus des radiofréquences, des protocoles de communication, du comportement des objets en activité et des canaux cachés.

8.1.1 Identification par radiofréquence

L'identification et la classification des émetteurs par l'intermédiaire de leur communication radiofréquence est un domaine scientifique exploré depuis quelques années. Nous pouvons citer les travaux réalisés sur le RFID [168, 169]. Il connait un regain d'intérêt avec le développement des objets connectés. Le but recherché est d'identifier avec précision un équipement mettant en évidence des caractéristiques uniques du champ électromagnétique émis [170]. Ces études requièrent généralement l'utilisation de matériels de laboratoire coûteux. Ainsi, des émetteurs 3G (*Universal Mobile Telephone System*) se distinguent avec plus de 99% de réussite [171]. Cette expérimentation laboratoire est reproductible dans un environnement complexe *indoor* avec de nombreux multi-trajets. Une étude similaire menée sur la norme *Institute of Electrical and Electronics Engineers* (IEEE) 802.11a a montré que la déviation par rapport à la fréquence centrale constitue un indicateur pertinent [172]. Récemment, l'arrivée des radio-logicielles a permis d'envisager d'exécuter cette opération d'identification à bas coût dans un environnement réel sur la norme IEEE 802.11a [173] et sur la norme IEEE 802.15 [5]. Ces travaux retournent une distinction précise de *smartphones* au regard de leurs émissions (cf. Figure 8.1). Ces données électromagnétiques de la couche physique du modèle *Open Systems Interconnection* (OSI) peuvent être enrichies avec les données des couches supérieures notamment de la couche MAC. Cette solution améliore leur assimilation. En-effet, les implémentations de la couche MAC sont généralement propriétaires [174]. Cette technique d'empreinte est également utilisée dans le cadre d'une détection de drones [175], en exploitant par exemple les variations électromagnétiques dues à la rotation des hélices de l'appareil. La question se pose réside dans la généralisation et l'industrialisation de cette démarche d'identification à d'autres protocoles de communication tels que LoRa ou SigFox notamment en l'adaptation des algorithmes. Est-ce que le critère de déviation par rapport à la fréquence centrale constitue un indicateur pertinent et universel à tous les protocoles ? Est-il possible d'envisager passivement une caractérisation matérielle précise d'un objet connecté ? Quel est l'impact de l'environnement sur la mesure ?

L'identification des objets communicants sert à détecter des comportements malveillants cherchant à masquer leur identité en modifiant volontairement leurs données [176]. Néanmoins, ce comportement est uniquement visible en simulation. Les variations d'empreintes

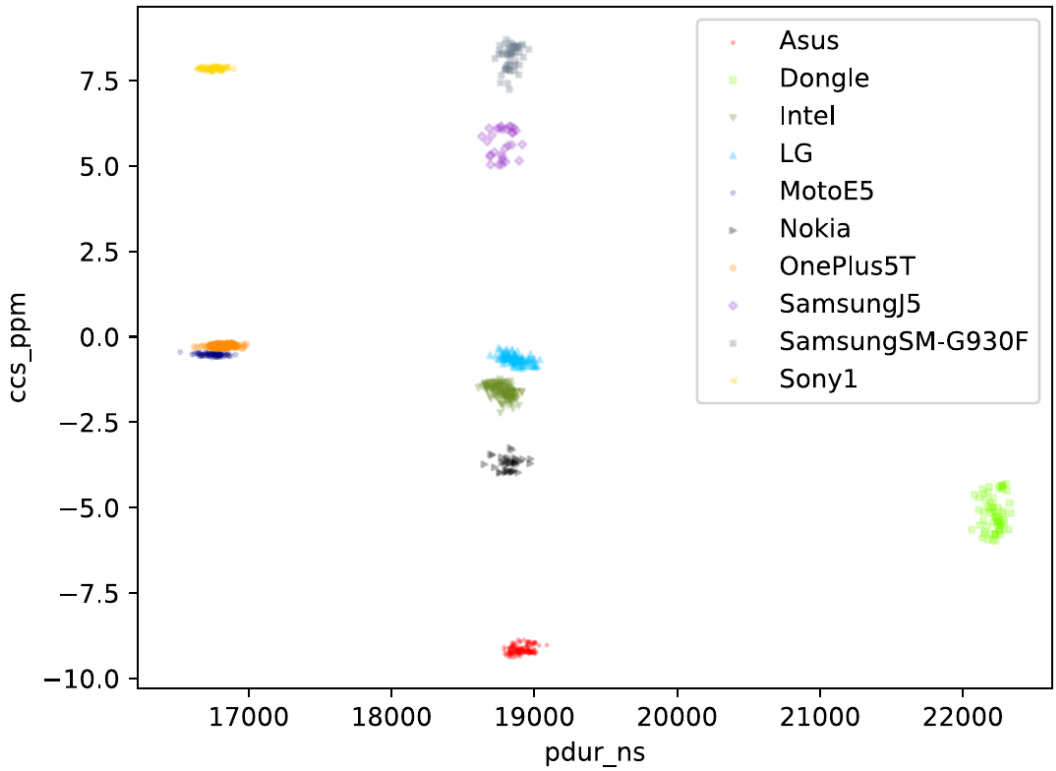


FIGURE 8.1 – Communications Bluetooth classées en fonction de la durée de préambule variable par rapport au décalage de l'horloge de la porteuse CCS - Source : [5]

sont subtiles. Elles sont drastiquement affectées par de nombreux facteurs environnementaux tels que la température ou le rapport signal à bruit [177]. En utilisant un principe similaire mais en champ proche avec une antenne large bande à proximité d'un microcontrôleur, les modifications du signal lors de l'exécution des instructions sont détectables [178]. Les études avec une solution radio-logicielle ont récemment montré que les empreintes en champ proche sont en mesure de conduire à la détection de programmes malicieux [179].

8.1.2 Identification protocolaire

Dans la littérature, les approches « informatiques » de l'identification vont principalement s'intéresser aux couches 2 à 7 du modèle OSI. Elles s'inspirent de travaux existants, comme ceux portant sur la détection du système d'exploitation d'un ordinateur telle que la discrimination d'un dispositif en fonction des ports ouverts ou des champs des entêtes IP. L'étude des ports ouverts est intéressante et généralisable notamment pour caractériser les objets locaux disposant d'une interface Ethernet. Le chapitre 4 aborde ce processus d'investigation dans le développement d'un outil d'identification rapide couplant des données lisibles de l'objet et un

module de pré-analyse réseau. Ce travail couvre une partie du spectre des équipements de l'Internet des objets tels que des passerelles ou les nœuds du réseau. Ces informations sont renseignées dans la base technique de connaissances des dispositifs connectés.

Plusieurs travaux scientifiques traitent le sujet de l'identification fine des dispositifs connectés par l'analyse des signaux. La plate-forme IoTScanner [180] détermine la présence et la position des objets connectés dans l'infrastructure afin notamment de détecter une attaque. Cette solution opère sur les protocoles Wi-Fi *Bluetooth Low Energy* (BLE) et Zigbee à partir de radio-logicielles à bas coût. Elle écoute de manière passive les transmissions radiofréquences sans connaissance préalable d'éléments de configuration, tels que le *Service Set IDentifier* (SSID) ou les clés de chiffrements. Les principales informations récupérées sont les types de trame, leur taille et leur adresse MAC, offrant une identification de l'objet connecté. Cette approche s'appuie très fortement sur des caractéristiques de protocole. Cette démarche ne semble pas généralisable pour des protocoles réalisés de manière *ad hoc*. Elle n'autorise pas le regroupement d'objets de même type. De nombreux autres travaux se sont également intéressés au Wi-Fi en étudiant les dérives d'horloge et les temps de réponse. La question est de savoir si ces solutions sont exploitables pour tous les types d'équipement connecté en particulier face à des solutions propriétaires.

8.1.3 Identification fondée sur l'activité de l'objet

IoTSense [181] propose de réaliser une empreinte numérique en se travaillant sur le comportement des objets, c'est-à-dire, la suite d'opérations réalisée par l'objet connecté lors d'une session de communication (*envoi E1* $\rightarrow$ *réponse R3* $\rightarrow$ *envoi E2* $\rightarrow$ *réponse R1 etc.*). Cette configuration s'appuie sur le postulat que les communications des objets connectés sont souvent courtes avec un temps entre deux sessions relativement long. Par exemple, la remontée d'une sonde de température *XIAOMI Mijia LYWSD03MMC* dans son fonctionnement nominal n'a lieu que toutes les 10 minutes. Cette méthode est pertinente pour notre problématique d'identification. Néanmoins, elle s'appuie uniquement sur l'étude de protocoles classiques de type *Hypertext Transfer Protocol* (HTTP) et TCP, peu utilisés pour les objets connectés de faible puissance. Elle doit donc être adaptée à notre problématique et généralisée sur les protocoles de communication sans-fil. N. Aluthge [182] propose d'étudier les séquences de démarrage des objets connectés afin de les identifier et de les classer. Cependant, cette démarche n'est pas toujours généralisable dans le cadre de l'investigation criminelle. En-effet, les matériels étudiés sont composés de mémoires volatiles RAM de type *Static Access Memory* (SRAM), *Dynamic Random Access Memory* (DRAM) ou *Non Volatile Random Access Memory* (NVRAM) et de mémoires non-volatiles *Read Only Memory* (ROM) de type *Mask ROM* (MROM) et *Programmable ROM* (PROM). L'absence d'alimentation d'une mémoire volatile risque d'entraîner une

perte d'information. La SRAM [183] et la DRAM [184] conservent les informations stockées dessus tant qu'une tension est appliquée entre les bornes.

8.1.4 Identification au travers des canaux cachés

Un certain nombre de travaux de la communauté scientifique démontre la faillibilité de tout système informatique face à des attaques physiques. En effet, l'activité même d'un microcontrôleur génère un ensemble de phénomènes physiques quantifiables, observables et exploitables. En ce sens, le monde physique constitue un « canal caché » de premier plan, en mesure de déterminer une information stockée et traitée. Parmi les travaux les plus pertinents sur ce sujet, trois familles d'attaques sont documentées par la communauté scientifique.

La première de ces trois familles regroupe l'ensemble des attaques de composants informatiques reposant sur l'analyse de la consommation énergétique aussi appelé *Differential Power Analysis* (DPA). Ces travaux étaient initialement un moyen de contourner la sécurité des systèmes de carte à puce [185]. Il s'agit alors de capturer une série de traces de la consommation énergétique du matériel à attaquer. Il est alors possible de corréler les variations de consommation énergétique et l'activité du matériel, afin d'obtenir une signature typique pour chaque opération et pour chaque valeur manipulée par le matériel. Il devient alors possible d'inférer la nature des traitements opérés en considérant les traces énergétiques dans leur ensemble. La valeur des données associées à ces traces est alors déterminée. La mise en œuvre de ces techniques est fastidieuse. Elle suppose une connaissance préalable d'un ensemble conséquent d'informations sur le matériel à attaquer. Ces techniques ont depuis été adaptées à un large spectre de matériel, et leur mise en œuvre a été largement simplifiée/optimisée par des procédés statistiques issus de l'étude du traitement du signal [186]. En facilitant les techniques de DPA, la communauté scientifique ouvre des perspectives d'industrialisation de l'approche. Néanmoins, les prérequis dans la connaissance du matériel à attaquer restent un verrou technologique important pour l'utilisation intensive de cette approche appliquée à un large spectre de système connecté de l'Internet des objets. En effet, les matériels exploités sont hétérogènes et variables d'une série à la suivante. Cependant, dans [187], les auteurs proposent une solution pour déterminer de manière semi-automatique, la signature énergétique des composants impliqués dans les matériels de l'IdO et des systèmes embarqués en général. Cette technique a été étudiée afin d'aider les concepteurs de logiciels embarqués dans la prédiction de l'énergie consommée par leur système. Toutefois, elle nécessite un accès physique à l'électronique et donc un démontage du matériel à analyser.

La seconde famille d'attaque regroupe les attaques relatives au rayonnement électromagnétique du support informatique, embarqué dans les objets connectés. Il s'agit, soit passivement d'écouter le rayonnement électromagnétique du composant en activité, soit activement de

générer un rayonnement électromagnétique vers le composant étudié afin d'induire un dysfonctionnement de ce dernier. Dans le premier cas, il est démontré que l'écoute du rayonnement électromagnétique permet d'observer le même type de phénomène qu'obtenu lors de la DPA. Selon le traitement effectué, un composant rayonne un signal différent. On parle alors d'analyse électromagnétique ou *Electromagnetic Analysis* (EMA). En analysant ces signaux, il est possible d'extraire la donnée « secrète » embarquée dans le composant [188, 189]. Cette attaque passive, par mesure de l'altération d'un signal connu, constitue une nouvelle famille d'attaque EMA ouvrant des perspectives dans l'extraction rapide de données contenues dans des objets connectés. Il est à noter qu'il ne s'agit pas ici d'injecter un signal suffisamment puissant pour engendrer des dysfonctionnements au sein du composant électronique, mais seulement d'en observer l'impact sur le signal transmis. Il s'agit bien d'une approche non intrusive. Dans le second cas, la mise en œuvre de l'attaque est active et intrusive, c'est-à-dire qu'elle impacte le fonctionnement du composant étudié. Il s'agit alors de générer des signaux puissants qui vont interférer avec le bon fonctionnement du matériel. La littérature scientifique parle d'injection de faute (*fault injection*) [190, 191, 192, 193, 194, 195, 196, 197]. Ce type de technique d'extraction de données manipulées par un système embarqué est comparable à des techniques d'escalade de privilège par exploitation d'une faille de sécurité informatique [198]. Ces techniques, souvent très efficaces, soulèvent cependant un nouveau verrou technologique lorsque nous souhaitons les appliquer à la criminalistique. Elles altèrent le support matériel. Dès lors la question de la pérennité des scellés et au-delà de la conservation de la preuve juridique est ouverte. Il s'agit de démontrer que l'opération qui a été faite par l'expert, souvent irréversible, a été suffisamment documentée et que les protocoles expérimentaux ont été suffisamment bien suivis pour que la preuve soit recevable devant une juridiction. La traçabilité garantit un suivi des actions dites « maîtrisées » ou « contrôlées » dans le cadre de l'investigation.

Enfin, il existe une troisième famille d'attaque des composants matériels qui exploite des propriétés physiques moins évidentes que la DPA ou la EMA comme les « canaux auxiliaires » pour voler de l'information [199]. Ces propriétés sont simples à quantifier et s'avèrent de redoutables moyens d'extraire de la donnée « secrète » manipulée par des microcontrôleurs. Le premier canal caché est tout simplement le temps. En effet, le temps de réponse d'un objet connecté est une source d'information sur le fonctionnement du système. Là encore, le domaine de la carte à puce fut, historiquement, le premier à être victime de ce type d'attaque [200]. Cependant, il a été montré que de nombreuses informations pouvaient être déduites à propos d'un système d'exploitation, simplement en observant la taille des paquets transmis en réponse à une demande de connexion TCP [201] ou en prenant en compte la taille des paquets d'un trafic réseau de l'application *Skype* [202, 203]. Or, la taille d'un paquet est une grandeur directement corrélée au temps de transmission. Une simple et précise mesure de temps infère de l'information sur les logiciels embarqués. Cette approche est passive et non-intrusive par

essence. Là encore, si l'état de l'art démontre la faisabilité de l'extraction, il reste silencieux sur son industrialisation, dans un contexte de forte hétérogénéité des composants électroniques utilisés et de gros volumes d'extraction de données à mener. Dans cette famille d'attaque par canaux cachés, nous notons encore la présence de techniques basées sur l'analyse des signaux audiofréquences générés par le composant. En effet, il est démontré qu'il est possible d'inférer de l'information sur un système informatique simplement en analysant le bruit qu'il génère. Ce bruit est lié à l'agitation thermique du composant et trahit donc son activité. Au vu des résultats rapportés par [204], la question n'est plus « est-il possible d'extraire de l'information simplement par l'analyse du bruit ? » mais « est-il possible d'automatiser cette extraction avec des solutions à bas coût et de construire une base de données des signatures acoustiques des composants embarqués de l'Internet des objets ? ». En effet, le verrou qui n'est pas à ce jour traité par l'état de l'art est de déterminer s'il est possible de classifier automatiquement des composants matériels et logiciels par ce type d'information.

Les différentes approches présentées proposent plusieurs pistes de classification mais leur spectre d'utilisation est beaucoup trop étroit. Elles ne s'appliquent pas complétement à tous les scénarios que nous devons traiter dans le cadre d'affaires judiciaires. Afin d'apporter une solution satisfaisante et opérationnelle à un coût raisonnable, nous devons coupler les différentes approches logicielles et matérielles. Ce couplage de procédés d'identification ne se retrouve pas dans la littérature. Ces différents attributs enrichissent notre base technique de connaissance des équipements.

8.2 Caractéristiques sonores d'un équipement électronique connecté

Les émanations acoustiques constituent un canal exploitable dans l'identification des objets connectés. Cette solution est d'autant plus pertinente à la criminalistique numérique car elle n'implique pas d'interaction et/ou de modification du support physique. La mesure est réalisée en périphérie du matériel, sans connaissance préalable. Il existe de nombreuses sources sonores exploitables dans un environnement informatique telles qu'une écoute de claviers ou d'imprimantes [205, 206, 207], un bruit mécanique lié à des ventilateurs ou les balais de lecture d'un dispositif de stockage. Ces informations ne sont pas transposables aux équipements numériques composant l'Internet des objets. Notre étude porte sur l'analyse du bruit émis par des appareils électroniques occasionné lors du passage du courant électrique. Il réside en des vibrations de composants électroniques, parfois entendues sous la forme d'un faible son aigu ou un sifflement communément appelé « *coil whine* », bien que souvent générées par des condensateurs. Ces émanations acoustiques, généralement causées par des circuits de régu-

lation de tension, sont corrélées avec l'activité du système puisque les processeurs changent radicalement leur puissance absorbée en relation avec le type d'opérations effectuées [208, 209]. Cependant, la largeur de bande de ces signaux est très faible : jusqu'à 20 *Kilohertz* (kHz) pour les signaux audibles mesurables par des microphones classiques et quelques centaines de kHz en utilisant des microphones à ultrasons. Au-delà de ces fréquences, l'atténuation de l'air et la sensibilité réduite des microphones rendent les signaux indétectables. L'étude des signaux doit permettre de discriminer un appareil, de reconnaître son taux d'activité et de le caractériser. Les opérations des processeurs étant de l'ordre du GHz ne sont pas observables en l'état. Cette information est noyée dans le bruit à partir de 300 kHz.

8.3 Description du dispositif expérimental

Le dispositif de mesure est composé d'une chaîne d'acquisition et de traitement de la donnée sonore, déployé au sein d'un environnement contrôlé. Il s'appuie sur des solutions économiques et transportables dans le cadre d'une intervention sur le terrain.

8.3.1 Chaîne d'acquisition des mesures sonores et de traitement

Le cahier des charges de la chaîne d'acquisition des mesures sonores est conditionné par des besoins de robustesse, d'adaptabilité à l'environnement et de faible coût. À terme, elle est susceptible d'être déployée à grande échelle dans des unités élémentaires de terrain de la Gendarmerie Nationale. Pour l'expérimentation, nous capturons le bruit acoustique émanant de divers dispositifs communicants en utilisant un microphone *BEHRINGER ECM8000* (efficace jusqu'à environ 98 kHz, bien au-delà de sa plage nominale de 20 kHz). Cet équipement a une réponse en fréquence très linéaire. Pour l'alimentation électrique et l'amplification, il est connecté à un *STEINBERG UR12*. Le signal amplifié de cette chaîne d'acquisition est numérisé à un taux d'échantillonnage de 192 kHz (cf. Figure 8.2).

FIGURE 8.2 – Chaîne d'acquisition des mesures sonores

Le traitement du signal audio est réalisé à partir d'un *Rasberry Pi*. L'acquisition des mesures s'appuie sur la librairie *PyAudio* de Python. Nous visualisons les spectrogrammes de ces signaux à l'aide de la fonction *spectrogram* de la librairie *SciPy* et par le biais de scripts

personnalisés. Ces développements offrent une modification de la résolution des résultats telle que le temps alloué à chaque Transformation de Fourier Rapide *Fast Fourier Transform* (FFT) avec leur taux d'entrelacement, les plages de fréquence à analyser ou de valeurs à retourner.

8.3.2 Atténuation du bruit

Afin de s'affranchir d'une partie du bruit extérieur et d'absorber les ondes sonores, nous élaborons une boîte anéchoïque acoustique (cf. Figure 8.3). Cette solution reproduit les conditions de champ libre et absorbe l'écho pouvant perturber les mesures. D'une dimension de 560x560x310 mm, elle est revêtue de dièdres en mousse polymère. Cet environnement d'acquisition est également isolé des vibrations extérieures par un support adapté. L'atténuation de la boîte est caractérisée à partir de la mesure d'une source sonore externe, variable en fréquence (cf. Figure 8.4). Un script établit la différence entre les mesures acoustiques ambiantes et de l'environnement isolé.

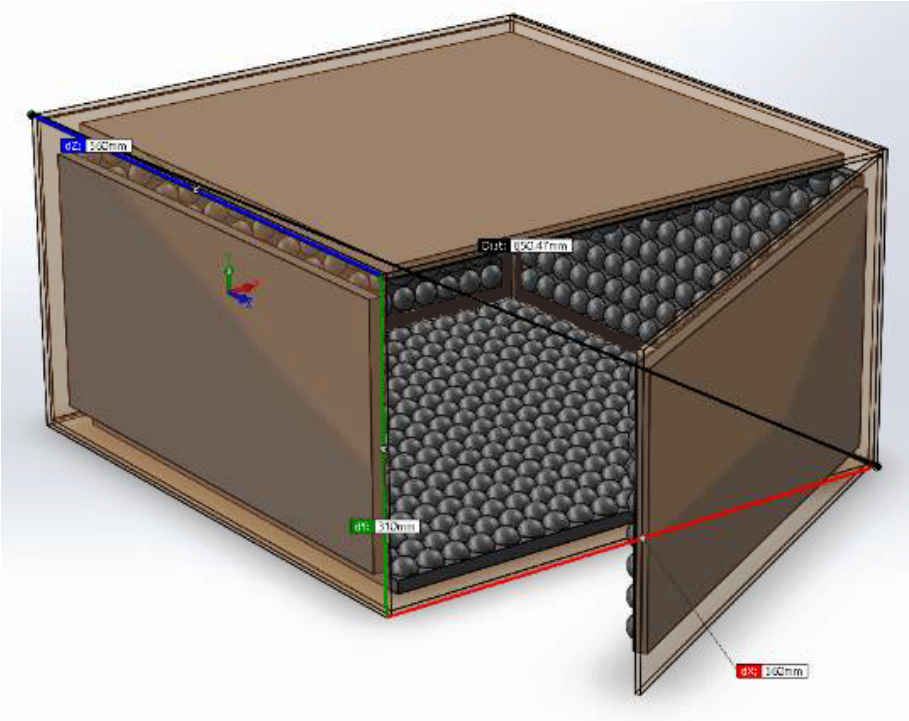


FIGURE 8.3 – Modélisation de la boîte anéchoïque acoustique

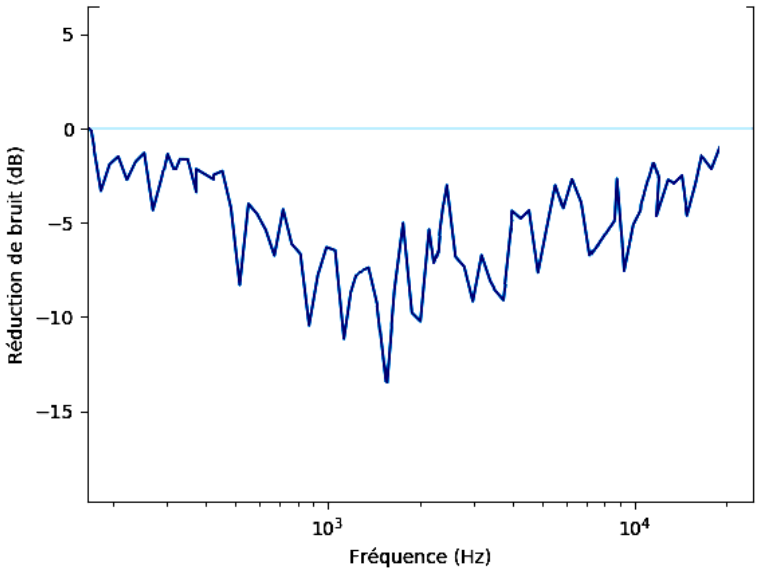


FIGURE 8.4 – Caractérisation en fréquence de l'isolation phonique de la boîte

La boîte anéchoïque acoustique contient uniquement le microphone de mesure. La chaîne d'acquisition est positionnée en extérieur afin de limiter la perturbation des mesures. Nous ajoutons un second microphone pour enregistrer le bruit ambiant. Cette donnée est soustraite de la mesure principale lors du traitement afin d'annuler les bruits ponctuels de l'extérieur (cf. Figure 8.5). Les graphiques du haut sont réalisés pour une mesure de son d'un chargeur USB et ceux du bas, dans le cadre d'une mesure de son d'une carte *STM32F769I-DISCO* de *STMicroelectronics*. Cette opération supprime une fréquence parasite autour de 64 kHz. Nous observons également une amélioration des fréquences basses et moyennes.

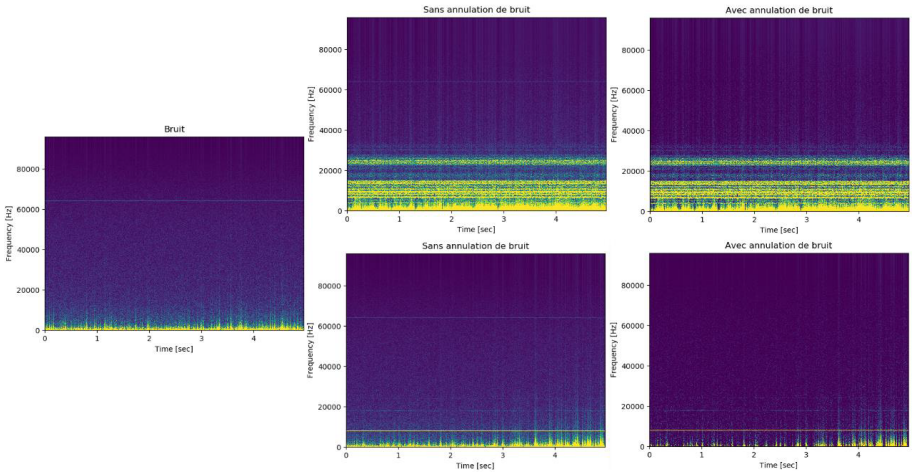


FIGURE 8.5 – Processus d'annulation du bruit ambiant

8.4 Analyse des fuites acoustiques d'équipements électroniques

Ces séries de mesures vérifient les propriétés acoustiques d'un équipement électronique. Elles cherchent à vérifier plusieurs hypothèses autour de la différenciation des matériels et de leurs comportements.

8.4.1 Discrimination des équipements connectés

Nous procédons à une série de mesure sur l'alimentation électrique de trois équipements de marques différentes : un *Motorola Moto G5 Plus* (téléphone n°1), un *Samsung Galaxy S4 mini* (téléphone n°2) et une carte *STM32F769I-DISCO*. Ce travail doit permettre de valider notre environnement de mesure et de vérifier l'hypothèse d'une différenciation des matériels branchés à partir du son émis lors d'une charge électrique. Pour cette expérience, nous utilisons un chargeur identique à chaque nouvelle évaluation de matériel. Les deux téléphones portables sont inactifs, avec le mode avion activé, le Bluetooth et Wi-Fi désactivés lors des enregistrements. Le mode avion est une configuration qui coupe toutes les connexions réseaux. Cependant, cette action laisse un accès libre aux applications et fonctions ne nécessitant pas d'être connectées. Nous constatons que le son émis par le chargeur autorise une différenciation des différents objets (cf. Figure 8.6). Nous observons également la présence d'une activité lors de la mesure sur les différents équipements, sans pouvoir la qualifier.

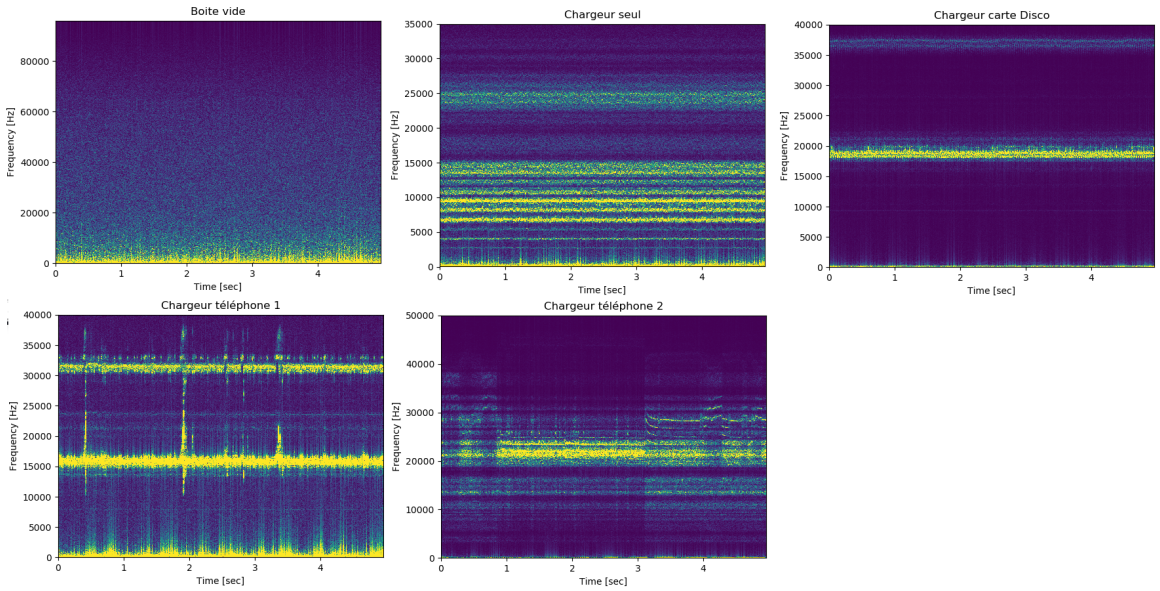


FIGURE 8.6 – Étude du son émis lors d'une charge électrique de dispositifs connectés

8.4.2 Qualification de l'activité d'un appareil

Par une seconde série de mesure sur le chargeur, nous cherchons à déterminer et à qualifier l'activité d'un appareil électronique (cf. Figure 8.7). Le téléphone utilisé pour la mesure est le *Motorola Moto G5 Plus* (téléphone n°1). Nous développons un programme sous Android réalisant l'incrémentation de 100 000 000 fois une variable et la recherche des 5 000 premiers nombres premiers. Ces opérations sont déclenchées à la suite d'un appui bouton. Nous opérons des mesures successives en mode avion activé, puis désactivé. Le lancement des opérations de calcul est mesuré sur les graphiques par un pic en fréquence, renseigné par un marquage en rouge sur le graphique. Nous remarquons une différence nette entre les phases d'activités lors des calculs et de repos. Il nous est difficile de caractériser précisément et visuellement un schéma type lié à une opération de calcul. Le processeur multi-cœur de notre téléphone est en mesure d'opérer plusieurs tâches en parallèle pouvant affecter le résultat de notre mesure. La désactivation du mode avion fait apparaître de nombreux pics en fréquence liés aux connexions réseaux.

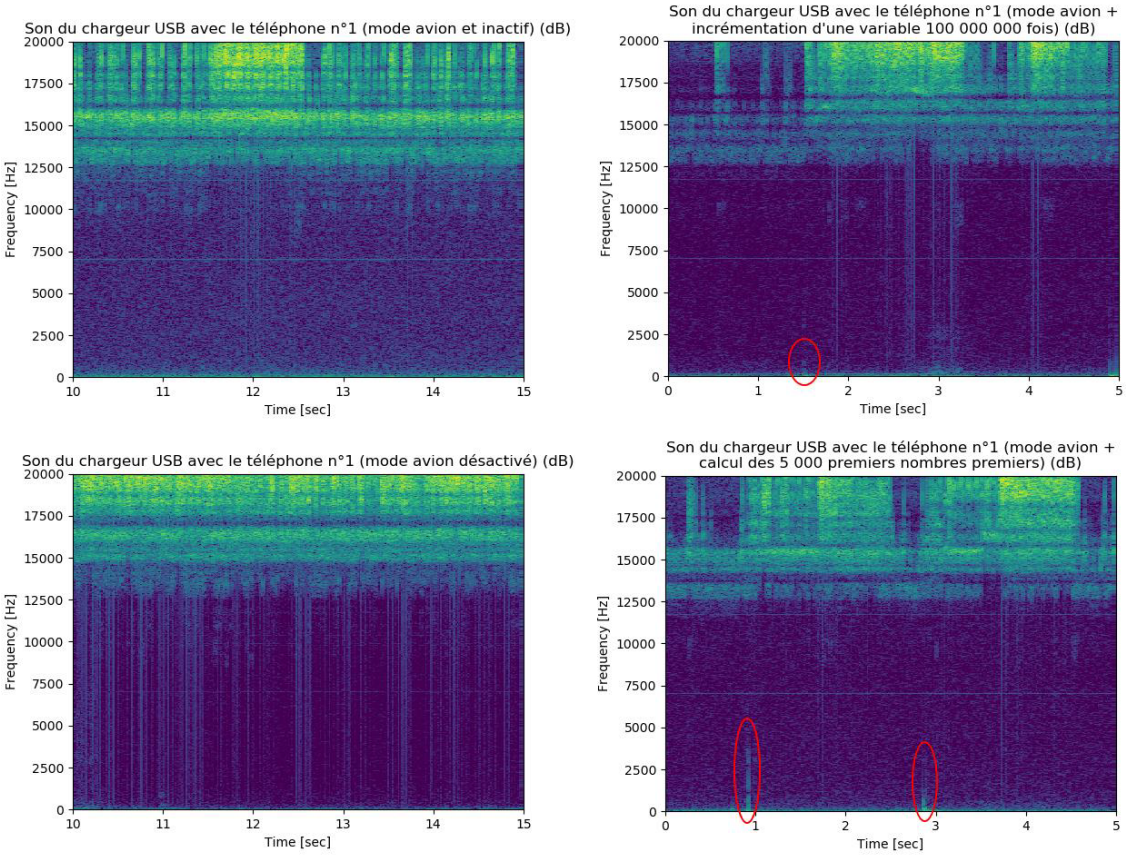


FIGURE 8.7 – Étude du son émis par un chargeur au regard de l'activité programmée sur un dispositif connecté

Par une troisième série de mesure, nous cherchons à déterminer et qualifier l'activité d'un appareil sans passer par le chargeur. Nos mesures sont opérées sur la carte *STM32F769I-DISCO* de *STMicroelectronics*. Ce kit contient un microcontrôleur *STM32F769NIH6* basé sur le cœur ARM Cortex-M7 32 bits. Plusieurs configurations d'écran sont évaluées : le mode accueil, le mode démonstration sans animation et le mode démonstration avec animation intégrant des calculs en temps réel. Le résultat retourné par notre chaîne d'acquisition ne permet pas de détecter un état particulier de l'objet 8.8). Ce résultat s'explique par les limites techniques des équipements de la chaîne d'acquisition. Pour les mêmes raisons, nous n'avons été en mesure de qualifier plus finement la différenciation de deux équipements ayant des configurations identiques (marque, modèle et système d'exploitation). Le matériel laboratoire proposé par Genkin et al. [210] ou une solution basée sur un vibromètre laser peuvent apporter un début de réponse à cette question.

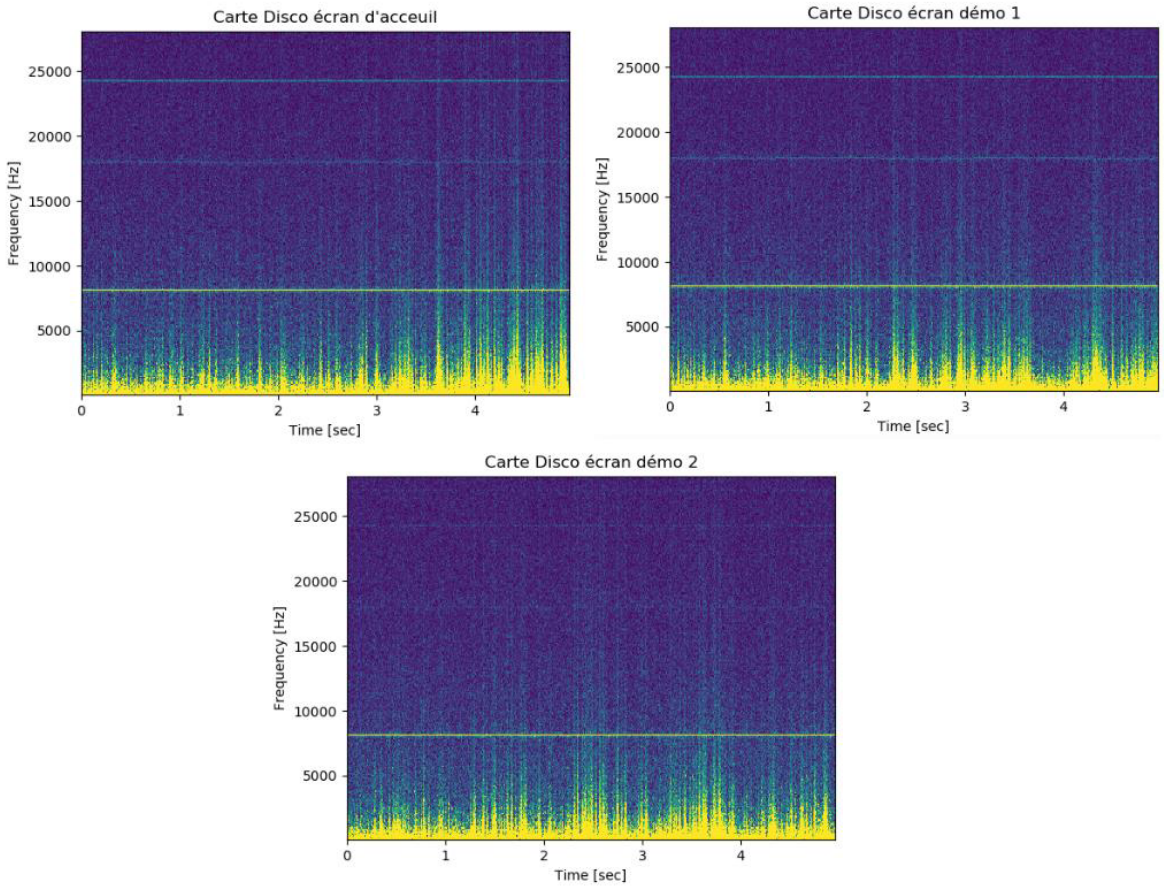


FIGURE 8.8 – Étude du son émis par un dispositif connecté

8.5 Automatisation de la classification des mesures

Les mesures acoustiques sur un même chargeur électrique lié à plusieurs équipements mettent en évidence des traces de signaux différents offrant la possibilité de les discriminer (mesure 1). Cependant, nous ne sommes pas en mesure de caractériser visuellement une activité sur un dispositif électronique (mesure 2 et 3). Dans cette section, nous cherchons à établir une classification automatique des équipements au regard de leurs signatures sonores. Pour ce point, nous utilisons une réduction par l'analyse en composantes principales PCA et une classification avec des machines à vecteurs de support *Support Vector Machine* (SVM).

8.5.1 Analyse en composantes principales

L'analyse en composantes principales (PCA) est une méthode descriptive utilisée pour explorer des données dites multi-variées, soit des données avec plusieurs variables. Elle donne une représentation graphique de l'information contenue dans un tableau de données quantitatives. Un tableau à n dimensions donne n composantes principales. Ces nouvelles variables correspondent à une combinaison linéaire des variables originelles. L'objectif de cette méthode est d'identifier les composantes principales autour desquelles la variation des données est maximale (cf. Figure 8.9).

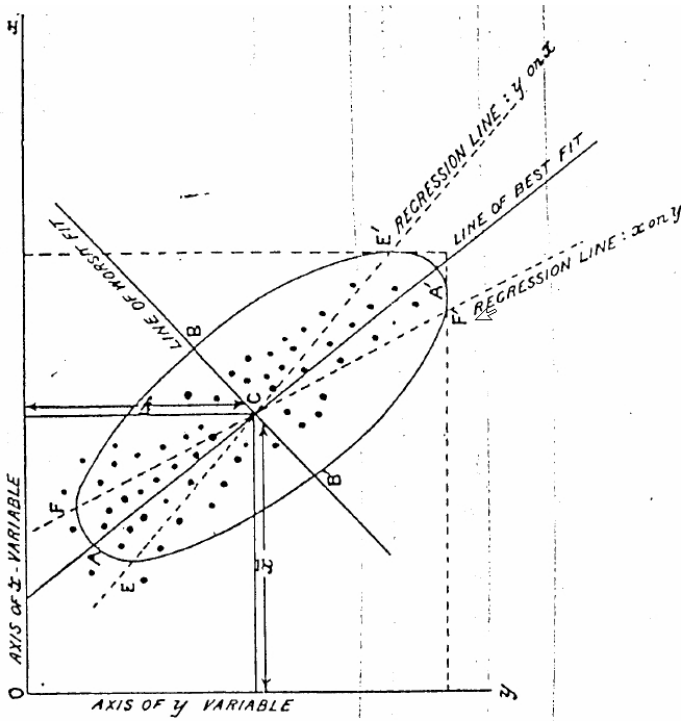


FIGURE 8.9 – Recherche de la droite du meilleur ajustement - Source : [6]

L'approche mathématique est modélisable selon le processus suivant (cf. Figure 8.10) :

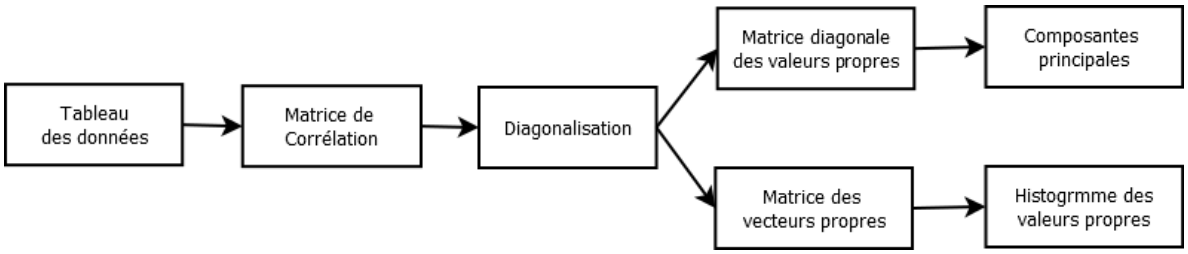


FIGURE 8.10 – Analyse en composantes principales

8.5.2 Machines à vecteurs de support

Les machines à vecteurs de support ou séparateurs à vaste marge SVM sont une classe d'algorithmes d'apprentissage, applicables à de la classification linéaire des données. Elles déterminent la frontière entre des catégories à partir des données d'entraînement. Après cette phase d'apprentissage, elles donnent une prédiction de la catégorie d'appartenance d'une entrée non rencontrée. Ce processus est opéré sans intervention humaine. Le développement de l'automatisation de la classification des mesures est réalisé en Python. Il s'appuie sur les librairies *numpy*, *sklearn*, *matplotlib*, et *customSpectrogram*.

8.5.3 Résultat de l'automatisation de la classification des mesures

Pour entraîner le SVM, nous effectuons des séries de mesures acoustiques sur un *Motorola Moto G5 Plus* sous Android 9 (*ROM : AOSPExtended Pie*) et un *Samsung Galaxy S4 Mini* sous Android 6.0.1 (*ROM : CyanogenMod 13*). Nous développons une application Android répétant des opérations mathématiques sur une longue période : l'incrémentation d'une variable, la multiplication de 2x2 et la vérification d'une condition 1=1. L'itération des opérations est réalisée par une simple boucle *for*. Afin que le système Android ne considère pas cette boucle comme infinie et tue le processus, nous intégrons dans le calcul des temps de pause périodiques. La phase d'apprentissage mène à une classification de 87 fichiers audio en 8 catégories.

Les graphiques (cf. Figure 8.11) montrent la PCA sans SVM et avec SVM en fonction de plusieurs noyaux : linéaire, radial et polynomiale. La frontière entre des catégories est affinable en augmentant le nombre de données d'entrainement. La méthode linéaire moins gourmande en ressource et en temps de calcul est à privilégier.

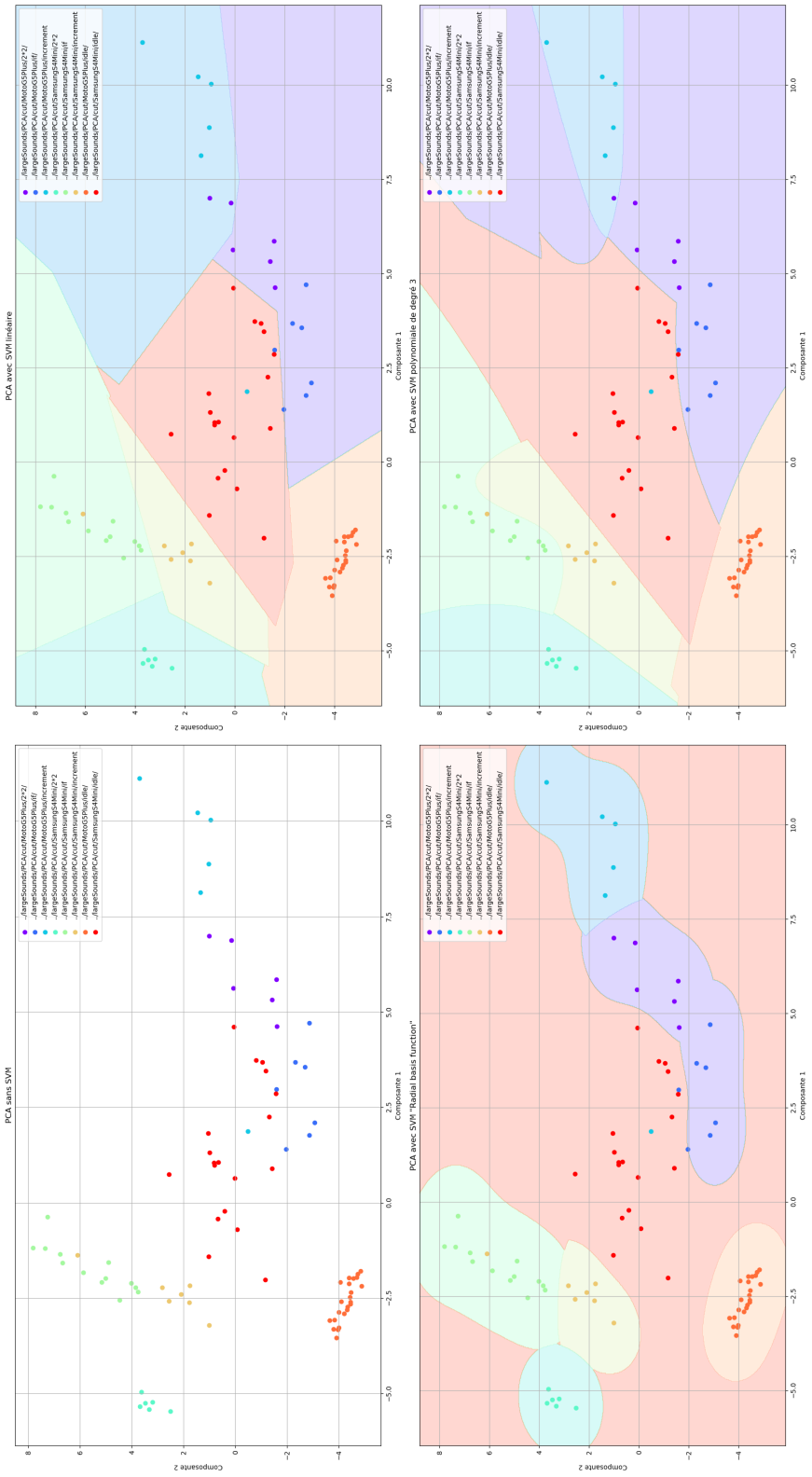


FIGURE 8.11 – Catégorisation par PCA

8.6 En quelques mots : identification fine d'un équipement connecté

Le domaine des objets connectés évolue très rapidement. Les recherches concernant la réalisation d'empreinte numérique des objets n'en est qu'à ses débuts. La majorité des approches existantes vont tenter d'identifier un objet en se basant sur un faible ensemble de caractéristiques de l'objet. Souvent ils ne s'intéressent qu'aux aspects matériels ou aux aspects logiciels, sans jamais confronter simultanément ces deux attributs, pourtant indissociables dans le cadre de l'Internet des objets.

L'étude acoustique des équipements électroniques donne des résultats intéressants dans une classification automatique des dispositifs et la constitution d'une base technique de connaissance des signatures acoustiques des composants. Cette opération est d'autant plus pertinente car elle est réalisée avec des outils à un faible coût. Elle permet une différentiation du matériel et dans certains cas un début de caractérisation fine de l'activité d'un appareil.

Troisième partie

Conclusion et perspectives

Chapitre 9

Conclusion et perspectives

"La science n'atteint jamais son but parce que le but n'en finit pas de se dérober
- et qu'en vérité il n'y a pas de but : la science est une tâche infinie.
Sa grandeur est de se présenter comme un rêve toujours inassouvi."
Jean d'Ormesson

9.1 Résumé des contributions

La massification des objets connectés entraîne une digitalisation du réel. Le numérique se superpose et interagit avec le monde physique. L'Internet des objets apporte une structuration de cet espace sous la forme d'une infrastructure cohérente. Elle se compose de groupes d'objets connectés (capteurs et actionneurs) mais également des passerelles, des interfaces de contrôle-commande et des plates-formes de centralisation et de traitement de la donnée. L'Internet des objets tend à s'organiser autour de services communs. Ainsi, l'assistant personnel virtuel est présent dans la domotique de la maison (compteurs intelligents, ampoules connectées, interphones, etc.), dans les équipements domestiques (télévision, box Internet, robots ménagers, enceintes connectées etc.) mais également dans les montres connectées et le multimédia des véhicules. Ce service commun façonne des écosystèmes cohérents en reliant des matériels hétérogènes par une caractéristique logicielle commune. Il crée par la même occasion une dépendance de solutions sous la forme de liens cachés. Cette structuration plurielle impacte directement les politiques d'échange et de remontée de l'information dans le réseau. Ce marché dynamique par sa quête constante d'innovation constitue une opportunité pour

l'investigation criminelle mais également un tournant dans la façon d'appréhender et de traiter l'espace numérique devenu complexe par les nombreuses interdépendances. L'objet connecté s'intègre et appartient à un tout. Il ne peut plus être traité de façon unitaire et doit donc être regardé selon l'infrastructure dès son appréhension initiale jusqu'à son analyse médico-légale.

Localement, la scène de crime est composée de nombreux dispositifs hétérogènes scrutant, interrogeant et analysant l'environnement direct. Ils génèrent une grande diversité de données liées aux utilisateurs, au système et à son fonctionnement. L'identification de ces éléments catalyseurs est primordiale pour appréhender avec efficacité ce milieu numérique : comprendre l'articulation de l'architecture connectée et collecter de l'information avec pertinence et efficience. Or, la diversité et la complexité des objets rendent jusqu'à présent la phase de reconnaissance relativement longue et disparate. À ce constat se joint l'agrégation des objets dans l'écosystème connecté avec des solutions de plus en plus transparentes et visibles pour les utilisateurs et les enquêteurs. Par nos travaux de thèse, nous étudions cette surface locale en cherchant à identifier tous les acteurs numériques. Nous proposons une articulation de l'identification en quatre étapes successives et itératives : une détection, une localisation, une reconnaissance des objets et un recoupement de l'information. Ainsi, nous abordons nos recherches par l'analyse des signatures radioélectriques des équipements. Les solutions techniques proposées dans la détection et la localisation des objets connectés s'appuient sur le RSSI et le déphasage des signaux émis. Afin de répondre à des contraintes opérationnelles et aux caractéristiques physiques de l'environnement, nous développons deux outils de captation : l'un, couvrant des surfaces importantes en ciblant des protocoles connus et l'autre, cartographiant avec minutie une zone délimitée sur une vaste gamme de fréquences et de protocoles. Le processus de reconnaissance des équipements est effectué à partir des informations visibles sur le produit et des signatures capturées sur la scène de crime. Nous développons un outil d'identification et de pré-analyse basé sur une base de données techniques des équipements. Une phase de recoupement de l'information finalise le processus d'identification. Elle s'appuie sur l'analyse des rôles des matériels présents sur la scène de crime. À l'issue de ce travail de recherche, il est intéressant d'étudier la question de la collecte et de la préservation des preuves contenues dans les équipements identifiés. Cette manipulation raisonnée constitue un défi majeur face à une infrastructure interconnectée et interdépendante. Cette structuration est sujette à des politiques spécifiques de gestion de l'information, tant au niveau de sa remontée dans le réseau que dans son stockage. Nous proposons une méthodologie de collecte en caractérisant l'infrastructure locale afin d'isoler efficacement les différents environnements. Les équipements sont collectés et conditionnés au regard des caractéristiques des objets, des données recherchées et des connaissances de la criminalistique numérique.

Les traces dans l'Internet des objets sont dispersées sur plusieurs niveaux : dans les équipements connectés, dans les ramifications de l'infrastructure et dans des espaces de traitement

en ligne. Les objets connectés sont physiquement présents sur la scène de crime ou rattachés à cette dernière par un lien qu'il soit caché ou non. Cette détermination de la présence et du positionnement de l'information demeure unique à chaque écosystème. Elle nécessite un travail de cartographie et de recherche des attributs communs. L'analyse des traces est donc beaucoup plus complexe que dans le cadre de la criminalistique numérique traditionnelle, en raison du caractère multidimensionnel et pluridisciplinaire de la donnée. Les dispositifs physiques sont traités en cohérence des connaissances en extraction et en décodage des systèmes électroniques par les sciences forensiques. Il en est de-même des données collectées dans le Cloud. Nos travaux de thèse réalisent un focus sur la problématique de l'information fragmentaire et dispersée. Nous développons une approche analytique de la donnée selon trois dimensions : le temps, l'espace et le contexte de l'événement au regard des rôles et des actions des équipements face au phénomène. Nous étudions le cycle de vie de la donnée dans l'infrastructure afin de la qualifier et d'établir sa cohérence. Le résultat réside en une chronologie de l'événement, ainsi que la détermination de ses causes et ses conséquences. Nous partons du postulat qu'il y a une dépendance de causes entre les différents événements et l'état de la donnée dans le système.

Dans une approche prospective, nous approfondissons les questions de l'identification fine des équipements connectés caractérisés par des critères objectifs et du développement de la base technique de connaissance pour les enquêteurs. Cette recherche part du postulat qu'il n'est pas possible de connaître à l'avance le fonctionnement de tous les objets disponibles sur le marché en plein essor et que l'enquêteur a besoin d'outils complémentaires pour identifier fidèlement les matériels numériques. En-effet, la reconnaissance visuelle des équipements admet des limites en l'absence de signes distinctifs. Le domaine des objets connectés évoluant très rapidement de façon hétérogène, les recherches concernant la réalisation d'une signature numérique des objets ou d'une famille d'objets n'en sont qu'à leurs débuts. La majorité des approches existantes tentent une identification fine en se basant sur un faible ensemble de spécificités techniques par rapport à un domaine d'étude restreint. Souvent, les recherches ne s'intéressent qu'aux aspects matériels ou aux aspects logiciels, sans jamais confronter simultanément ces deux aspects, pourtant indissociables dans le cadre de l'Internet des objets. De plus, de nombreuses méthodes proposées impliquent des interactions et/ou modification du support physique. Nous présentons donc une étude acoustique en périphérie des équipements électroniques par l'usage d'une chaîne d'acquisition à faible coût. À partir d'une démarche d'apprentissage, nous obtenons un début de caractérisation fine des équipements. Ce nouveau critère objectif complète notre palette dans l'identification des objets connectés.

9.2 Perspectives

Les travaux sur l'appréhension de la scène de crime contenant des dispositifs connectés soulèvent plusieurs observations sur le modèle forensique et sur la réponse criminalistique numérique de proximité. La formation et les moyens à disposition des unités opérationnelles sont susceptibles de nécessiter des adaptations à cette évolution technique. Les primo-intervenants sont souvent des enquêteurs d'unités élémentaires, non formés aux technologies avancées. Leur sensibilisation sur l'investigation dans le numérique peut être renforcée par un volet sur la reconnaissance visuelle des équipements connectés. Cette démarche d'identification doit être couplée avec les solutions existantes dans l'aide et l'assistance sur l'investigation numérique (portail *CyberAide* et *Guichet Unique Téléphonie Internet* (GUTI)). Une mallette criminalistique numérique projetable sur le terrain pourrait voir le jour sur un format similaire aux mallettes *Police Technique et Scientifique* (PTS) des *Technicien d'Identification Criminelle* (TIC) de proximité, existantes dans toutes unités élémentaires. Elle serait en mesure de regrouper les outils de détections, de localisation et d'identification rapide déclinés dans cette thèse avec des solutions spécifiques pour la collecte (sac de Faraday et batteries externes). En laboratoire, un environnement d'identification fine et d'analyse des interactions pourrait être utilement envisagé. Il s'intéresserait également à la reconstruction graphique de l'infrastructure numérique de la scène de crime, à sa virtualisation et à l'étude des rapprochements entre les écosystèmes connectés. À ces besoins opérationnels doit s'ajouter une réflexion menée sur l'emploi du technicien en nouvelles technologies dans la démarche d'investigation sur la scène au plus proche de l'événement criminel dans l'étude des signaux radioélectriques et le suivi de l'évolution des équipements connectés.

Le volet Cloud est en mesure d'orienter et d'optimiser la recherche des objets à saisir lors d'une perquisition ou dans l'étude d'une scène criminelle. En réponse à une réquisition judiciaire en amont des investigations, les opérateurs de services sont susceptibles d'identifier des équipements connectés. Cette recherche s'appuie sur les logs de connexion des appareils à une passerelle fixe. Les informations recueillies se présentent sous la forme du nom de l'appareil (par défaut ou modifié par l'utilisateur) et de l'adresse MAC de la carte réseau (MAC WiFi, MAC filaire Ethernet). Ces éléments renseignent l'ensemble des dispositifs connectés à la passerelle tels que les décodeurs d'une télévision, les ordinateurs, les *smartphones*, les assistants personnels virtuels, les systèmes domotiques, etc. Une fois la marque de l'appareil identifié, le fabricant peut-être en mesure de donner d'autres éléments à partir de l'adresse MAC comme une *International Mobile Equipment Identity* (IMEI), des informations de compte Cloud, des données vidéos ou photos, etc. L'IMEI permet de requérir les opérateurs téléphoniques pour obtenir toutes les informations liées à l'abonné mobile telles qu'une identité, des numéros de téléphone, des cartes SIM insérées, des détails de trafic, etc. Ces informations viennent en

complément des éléments présentés dans ces travaux de thèse et peuvent apporter des éléments décisifs pour l'enquête. Cependant, elles nécessitent une collaboration entre les opérateurs et les forces de sécurité.

Le besoin d'identification fine prend de l'importance avec le développement de solution « fait-maison » et les solutions hybrides. Dernièrement, la société STMicroelectronics propose un matériel basé sur un STM32H743VIT6E. Cet équipement embarque l'assistant personnel intelligent *Alexa* développé par la société *Amazon*. Ce matériel est plus flexible et polyvalent en termes d'usage et de configuration. Il est donc plus difficile à appréhender, à identifier et à analyser pour les enquêteurs, notamment en l'absence d'éléments visuels et de logiciels spécifiques à l'utilisateur. Dans certains cas, il est susceptible d'émettre sur des fréquences non autorisées. De plus, il peut se retrouver maquillé dans les objets du quotidien (peluche, coque d'*Amazon Echo*, boîte électrique, objets de décoration, etc.). Ainsi, le besoin de coupler les approches matériels et logiciels avec l'intelligence artificielle pour une identification fine devient primordiale afin d'apporter une réponse concrète à ces problématiques.

Par l'Internet des objets, la société se numérise et créée parallèlement un miroir de données. La trace de l'action criminelle singulière par nature se retrouve alors emprisonnée et déportée dans le cyberespace. À quand le crime parfait ?

Annexes
Exemple de forensique sur certains objets connectés

Annexe A

Kit domotique Orvibo (5)

A.1 Passerelle Orvibo

Version du hub : 2.3.0
Système d'exploitation : Linux DD-WRT 2.6.24

A.1.1 Structure du système d'exploitation

Répertoire	**Utilisation**
/bin	Stockage des binaires (initialisation système et commandes principales)
/dev	Stockage des fichiers servant de communication avec les périphériques
/etc	Stockage des fichiers de configuration et des scripts de paramétrage
/home	Racine des répertoires utilisateurs
/lib	Stockage des bibliothèques du noyau
/mnt	Racine des points de montage des autres systèmes de fichiers
/proc	Système de fichier virtuel, "image" du système
/sys	Similaire à */proc*, présent sur les anciennes version de Linux
/usr	Stockage des données pouvant être partagées entre utilisateurs
/var	Stockage des données fréquemment écrites

A.1.2 Données liées à la passerelle Orvibo

Fichier	**Répertoire du fichier**	**Éléments exploitables**
version		Version du hub
TZ	*/mnt/ProgramFiles/etc/*	Fuseau horaire (TimeZone)
vihomed-config.json	*/mnt/ProgramFiles/usr/local/*	Numéro de modèle Interface réseau
zigbee-channel	*/mnt/*	Éléments de configuration ZigBee : adresse réseau et physique
kvdata.db	*/mnt/*	Configuration du hub : - timezone_offset - configuration de l'environnement utilisateur - familyId
vihome2.db	*/mnt/*	Configuration de l'utilisateur : - Informations du compte - Informations de la famille Configuration du réseau ZigBee : - Informations du hub - Adresses réseau des objets associés au hub Configuration physique de l'environnement : - Détail des étages - Détail des pièces Configuration logique de l'environnement : - Détail des objets (hub, capteur, etc.) - Détail des regroupements d'objets par pièce - Scènes d'utilisation programmées - Scénarios de sécurité programmés Traces d'interactions avec l'environnement : - Horodatage de l'enregistrement des scènes et des scénarios - Horodatage des événements de l'environnement, comme le déclenchement d'un capteur, le changement d'une image à la caméra, etc.
log.db	*/mnt/*	Derniers événements survenus slon trois catégories (Info, Warn et Error)
run.log run.log.0	*/mnt/*	Informations liées aux activités du réseau ZigBee et au fonctionnement du hub (sur les dernières heures de fonctionnement)

A.2 Application Orvibo

Application Orvibo version 3.4.3 (/mobile/Applications/) installée le téléphone iPhone SE

A.2.1 com.orvibo.cloudPlatform

Répertoire du fichier	**Éléments exploitables**	**Apport**
/Document/.tencent_analysis _WXOMTAStore	Evènements issus des manipulations de l'utilisateur	Identification et horodatage des appairages aux points d'accès
/Document/HomeMate.db	Détail du compte Détail des étages Détail des pièces Détail des objets Détail des scènes Détail des scénarios Détails des événements Configuration du hub	Configuration physique de l'environnement (pièces) Configuration logique de l'environnement (matériel et regroupement d'objets) Identification des utilisateurs et d'informations réutilisables (mot de passe)
/Library/Preferences/com. orvibo.cloud/Platform.plist	Détail du compte Informations du point d'accès Géolocalisation de l'appareil	Identification de l'utilisateur et d'informations réutilisables Identification du dernier point d'accès appairé Identification du dernier emplacement connu par l'application

A.2.2 group.com.orvibo.HomeMateWidget

Répertoire du fichier	**Éléments exploitables**	**Apport**
/Library/Preferences/group .com.orvibo.HomeMateWidget .plist	Détail du compte Détail des modes de sécurité	Identification de l'utilisateur et d'informations réutilisables Identification des modes de sécurité programmés/actifs

Annexe B

Kit domotique Philips Hue (8)

B.1 Passerelle Philips

Version du hub : Philips Hue bridge 2.0
Système d'exploitation : Linux 3.14.0

Clé de l'API	**Éléments exploitables**	**Apport**
lights	Identifiant et nature des lampes	Configuration logique de l'environnement (matériel)
groups	Informations des groupes et listes des lampes	Configuration physique de l'environnement (pièces) Configuration logique de l'envrionnement (regroupement d'objets)
config	Informations du hub et white-list	Configuration du hub et état de la connexion au cloud Identification des utilisateurs et horodatage de la première et dernier utilisation par utilisateur
schedules	Détail des routines	Identification et horodatage d'une interaction avec l'utilisateur
scenes	Informations des scènes	Configuration logique del'environnement Identification et horodatage d'une interaction avc l'utilisateur
sensors	Identifiantsdes capteurs	Configuration logique de l'environnement (matériel)

B.2 Application Philips Hue

Application Philips Hue version 3.2.0 (/mobile/Applications/)

B.2.1 com.philips.lighting.hue2

Répertoire du fichier	**Éléments exploitables**	**Apport**
/Library/com.amplitude. database	Evènements de l'application	Identification et horodatage d'interactions avec l'environnement

B.2.2 group.com.philips.hue2

Répertoire du fichier	**Éléments exploitables**	**Apport**
/Hue.sqlite	Détail des objets Informations de pièces Informations des scènes	Configuration logique de l'environnement (matériel et regroupement d'objets) Identification et horodatage d'interactions avec l'environnement (configuration)
*/debuglog_**	Logs d'activités	Identification et horodatage d'interactions avec l'environnement
/Library/Preferences/group .com.philips.hue2.plist	Détail des hubs appairés	Configuration logique de l'environnement (matériel)

Annexe C

Amazon Echo (14)

C.1 Amazon Echo Dot

Système d'exploitation : Fire OS (fork d'Android fondé sur un noyau linux)

Fichier	**Répertoire du fichier**	**Éléments exploitables**
Build.prop (P13)(P14)		ro.build.version.name ro.build.version.release ro.build.date ro.product.model ro.product.manufacturer ro.product.cpu.abi ro.build.fingerprint
Settings.db(P16)	*/data/com.android.providers .settings/databases/* */data/com.amazon.providers .settings/databases/*	Device_name WiFi Country Code ATR Response Provider ID Android Nom d'appareil et adresse MAC Bluetooth Mot d'activation configuré (alexa_selected_wakework_model)
Persist.sys(P16)	*/property/*	persist.sys.wifi.country_code et persist.sys. country persist.sys.profiler_ms persist.sys.last_updated_build et persist.sys.last_ verified_build persist.sys.last_synced_time persist.sys.language
softap.conf(P16) wpa_supplicant.conf(P16) networkHistory.txt(P16)	*/misc/wifi/*	nom du WiFi DDID + BSSID

Fichier	**Répertoire du fichier**	**Éléments exploitables**
0.xml(P16) */0/accounts.db*(P16)	*/system/users/*	Informations sur le compte utilisateur (nom et type de service)
Logs	*/system/dropbox/*	Log.amazon_main_n Log.main_n Log_events_n Log.kernel_n Log_metrics_n Log.system_n Log.vitals_n

C.2 Application Amazon Echo

Application Amazon Echo 2.2.222061.0 (/mobile/ Applications/) installée le téléphone iPhone SE

Répertoire du fichier	**Éléments exploitables**	**Apport**
AlexaMobileiOSComms.sqlite	Historique des discussions	Discussions Contact associé (prénom, nom et numéro de téléphone) NB : historique mis à jour à partir du cloud à chaque réinstallation de l'application après synchronisation (messages antérieurs à l'installation retrouvables)
RCTAsyncLocalStorage_V1	Journal des cartes et historique des interactions vocales	Activité de l'utilisateur (format JSON) : par exemple question demandée par l'utilisateur datée URL pointant sur l'enregistrement sonore de la commande en ligne
/Library/Preferences/	Données utilisateur et Amazon-ID	Dernière date et l'heure de fermeture de l'application Version de l'application Nom et prénom de l'utilisateur Amazon-ID Date et l'heure d'importation des contacts etc.

Annexe D

Wink Hub 2 (17)

D.1 Passerelle Wink Hub 2

Version du hub : 2.0

Système d'exploitation : Linux 3.14.52

Répertoire		
/bin	*/lib*	*/opt*
/database	*/lib32*	*/proc*
/database_default	*/linuxrc*	*/root*
/dev	*/media*	*/run*
/etc	*/mfgtests*	*/sbin*
/home	*/mnt*	*/sys*
/tmp	*/usr*	*/var*

D.2 Application Wink

Application Wink sous Android version 7.0.18.23531 (/data/data/)

Fichier	**Répertoire du fichier**	**Eléments exploitables**
android.wink.wink_preferences.xml	*/com.quirky.android.wink .wink/shared_prefs/com.quirky*	Information utilisateur (adresse mail)
PersistanceDB	*/com.quirky.android.wink .wink/databases/*	Les objets associés et leurs configurations (nom, modèle, type d'objets, rôle, date de création, mise à jour, modification, dernière connexion, etc.)
user.xml	*/com.quirky.android.wink .wink/shared_prefs/*	Informations sur le compte utilisateur (nom et type de service)
wink_preferences.xml	*/com.quirky.android.wink .wink/shared_prefs/*	WiFi SSID et mot de passe (chiffré en Base64)
winkdevices.xml	*/com.quirky.android.wink .wink/shared_prefs/*	Activité et configuration des objets connectés au hub

Annexe E

Apple Watch 3 (18)

Système d'exploitation de la montre connectée : WatchOS
Liste des applications natives : Activity, Calendar, Clock, Contacts, Friends, Mail, Maps, Messages, Music, Passbook/Apple Pay, Phone, Photos, Reminders, Stocks, Weather, Workout.

Fichier	**Répertoire du fichier**	**Eléments exploitables**
Properties.bin	*/mobile/Library/Device Registry.state/*	Nom de la montre Modèle OS GUID Chemin de synchronisation des données
HistorySecureProperties.plist SecureProperties.bin	*/mobile/Library/Device Registry.state/*	Adresse MAC Wi-Fi Adresse MAC Bluetooth Numéro de série UDID de l'Apple Watch
StateMachine- <GUID>.plist	*/mobile/Library/Device Registry.state/*	Horodatage d'appariement Version de la montre lors de l'appariement
ActiveStateMachine.plist	*/mobile/Library/Device Registry.state/*	Version du système d'exploitation
Device Registry	*/mobile/Library/*	Registre
NanoDomains	*/mobile/Library/Device Registry/<GUID>/Nano Domains/com.apple.Carousel/*	Liste des applications installées
AddressBook	*/mobile/Library/Device Registry/<GUID>/Address Book/*	Favorites.previous

Fichier	**Répertoire du fichier**	**Eléments exploitables**
Nanoappregistry	*/mobile/Library/Device Registry/<GUID>/Nano AppRegistry/Applications/*	Applications installées sur la montre
registry.sqlite (application native NanoMail)	*/mobile/Library/Device Registry/<GUID>/Nano Mail/*	Compte de messagerie définis sur l'iPhone avec les adresse mails Horodatage de mails avec nom du dossier et adresses mails associées
Voicemails	*/mobile/Library/Device Registry/<GUID>/PreferencesSync/NanoDomains/com.apple.mobilephone/*	Numéro de téléphone et les chemins de synchronisation des fichiers voicemail
nanopasses.sqlite3 (application native Apple Wallet)	*/mobile/Library/Device* Registry/<GUID>/Nano passes/	Passbook : horodatage, nom, prix, adresses, etc.
Health Data (sauvegarde chiffrée)		Positions GPS Fréquences cardiaques Nombre de pas etc.

Acronymes

ADB *Android Debug Bridge*. 93

ADN *Acide Désoxyribonucléique*. 12, 79

AES *Advanced Encryption Standard*. 40

ANATEL *Agency of National TELecommunications*. 44

AoA *Angle of Arrival*. 53

API *Application Programming Interface*. 21, 72, 80, 85, 93

ARCEP *Autorité de Régulation des Communications Electroniques et des Postes*. 15

ARP *Address Resolution Protocol*. 46

ASCII *American Standard Code for Information Interchange*. 95

BLE *Bluetooth Low Energy*. 110

C3N *Centre de Lutte contre les Criminalités Numériques*. 64

CAN *Convertisseur Analogique-Numérique*. 40

CCS *Carrier Clock Skew*. viii, 109

CERT *Computer Emergency Response Team*. 15

CMIIT *China Minitry of Industry and Information Technology*. 44, 46, 49

CPP *Code de procédure pénale*. 40, 41, 79

CPS *Cyber-Physical System*. 14

dB *décibel*. 53

DNS *Domain Name System*. 103

DPA *Differential Power Analysis*. 111, 112

DRAM *Dynamic Random Access Memory*. 110, 111

EIRP *Effective Isotropic Radiated Power*. 64

EMA *Electromagnetic Analysis*. 112

FCC *Federal Communications Commission*. 44, 46, 49, 85

FFT *Fast Fourier Transform*. 115

FIB-SEM *Focused Ion Beam - Scanning Electron Microscope*. 93

FSWG *Cloud Computing Forensic Science Working Group*. 95

FTC *Federal Trade Commission*. 15

FTP *File Transfer Protocol*. 47

GHz *Gigahertz*. 49, 114

GPS *Global Positioning System*. 51, 59, 87

GSM *Global System for Mobile Communications*. 59, 61, 82

GUTI *Guichet Unique Téléphonie Internet*. 130

HTTP *Hypertext Transfer Protocol*. 110

IC *Industry Canada*. 44, 46

IDATE *Institut de l'audiovisuel et des télécommunications en Europe*. vii, 16

IDC *International Data Corporation*. 3

IdO *Internet des objets*. iii–v, vii, 3, 5, 8, 9, 13–16, 18, 19, 36–42, 51, 55, 56, 81, 83, 86, 87, 90, 91, 105, 111

IDS *Intrusion Detection System*. 25

IEC *International Electrotechnical Commission*. 15

IEEE *Institute of Electrical and Electronics Engineers*. 108

IHM *Interfaces Homme-Machine*. 17, 42, 47, 50, 73, 92

IMEI *International Mobile Equipment Identity*. 130

IoT *Internet of Things*. 3, 13

IoT-GSI *Internet of Things – Global Standards Initiative*. 14

IP *Internet Protocol*. 17, 46, 109

ISM *Industrielles, Scientifiques et Médicales*. 60

ISO *International Organization for Standardization*. 15

JSON *JavaScript Object Notation*. 100

JTAG *Joint Test Action Group*. 93

KCC/MSIP *Korean Communications Commission/ Ministry of Science, ICT and Future Planning*. 44, 46

kHz *Kilohertz*. 114, 116

Li-Fi *Light Fidelity*. 17, 57

LoRa *Long Range*. 17, 61, 76, 108

LoRaWAN *LoRa Wide Area Network*. 57, 78

MAC *Medium Access Control*. 41, 44–46, 57, 58, 103, 108, 110, 130

MHz *Mégahertz*. 47, 62, 64

MIT *Massachussetts Institute of Technology*. 13

MPU *Memory Protect Unit*. 92

MROM *Mask ROM*. 110

MVC *Modèle-vue-contrôleur*. vii, 46

mW *Milliwatt*. 49

NIST *National Institute of Standards and Technology*. 14, 33, 95

NTFS *New Technology File System*. 92

NVRAM *Non Volatile Random Access Memory*. 110

OS *Operating System*. 59, 93

OSI *Open Systems Interconnection*. 108, 109

PCA *Principal Component Analysis*. viii, 120–122

PROM *Programmable ROM*. 110

PTS *Police Technique et Scientifique*. 130

PVe *Procès-Verbal électronique*. 55

RAM *Random Access Memory*. 13, 40, 41, 92, 110

RFID *Radio Frequency Identification*. 14, 24, 108

ROM *Read Only Memory*. 110, 121

RPL *Routing Protocol for Low-Power and Lossy Networks*. 35, 56

RSSI *Received Signal Strength Indication*. 53, 54, 59, 68, 128

SD *Secure Digital*. 84, 93

SDK *Software Development Kit*. 90

SDR *Software Defined Radio*. 58, 62

SIM *Subscriber Identity/identification Module*. 77, 80, 82, 130

SRAM *Static Access Memory*. 110, 111

SSDP *Simple Service Discovery Protocol*. 56

SSH *Secure Shell*. 47

SSID *Service Set IDentifier*. 110

SVM *Support Vector Machine*. 120, 121

TAP *Test Access Port*. 41

TCP *Transmission Control Protocol*. 17, 40, 46, 110, 112

TDOA *Time Difference Of Arrival*. 53

Telnet *Terminal network ou Telecommunication network*. 47, 80

TIC *Technicien d'Identification Criminelle*. 130

UART *Universal Asynchronous Receiver Transmitter*. 93

UDP *User Datagram Protocol*. 40

UID *User IDentifier*. 44

UIT *Union International des Télécommunication*. 14

USB *Universal Serial Bus*. 93, 116

Wi-Fi *Wireless Fidelity*. 17, 26, 28, 53, 57, 59, 61, 62, 64, 68, 76, 78, 82–84, 110, 117

XML *Extensible Markup Language*. 100

Bibliographie

[1] DGE. Prospective - marchés des objets connectés à destination du grand public. https://www.entreprises.gouv.fr/files/files/directions_services/etudes-et-statistiques/prospective/Numerique/2018-05-24-Etude-objets-connectes.pdf.

[2] Nadir Cherifi. *Assistance au développement de logiciels embarqués contraints en énergie*. PhD thesis, Lille 1, 2018.

[3] Bruno Dorsemaine, Jean-Philippe Gaulier, Jean-Philippe Wary, Nizar Kheir, and Pascal Urien. A new approach to investigate iot threats based on a four layer model. In *2016 13th International Conference on New Technologies for Distributed Systems (NOTERE)*, pages 1–6. IEEE, 2016.

[4] Sundresan Perumal, Norita Md Norwawi, and Valliappan Raman. Internet of things (iot) digital forensic investigation model : Top-down forensic approach methodology. In *2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)*, pages 19–23. IEEE, 2015.

[5] Étienne Helluy-Lafont, Alexandre Boé, Gilles Grimaud, and Michaël Hauspie. Bluetooth devices fingerprinting using low cost sdr. In *2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC)*, pages 289–294. IEEE, 2020.

[6] Karl Pearson. Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(2) :559, 1901.

[7] IDC. Worldwide internet of things forecast, 2020–2024. https://www.idc.com/getdoc.jsp?containerId=US45861420, 2020.

[8] Statista. Estimation du marché des objets connectés en volume dans le monde en 2013, 2015 et 2020. https://fr.statista.com/statistiques/561092/volume-marche-objets-connectes-monde/, 2019.

[9] Statista. Objets connectés : valeur du marché mondial 2013-2020. https://fr.statista.com/themes/2972/les-objets-connectes/#dossierSummary__chapter1, 2020.

[10] Clémence Dunand. A hong kong, les manifestants adoptent firechat pour communiquer sans réseau. https://www.lesechos.fr/30/09/2014/lesechos.fr/0203817723785

_a-hong-kong--les-manifestants-adoptent-firechat-pour-communiquer-sans-reseau.htm, 2014.

[11] Éric Hazane. Sécurité numérique des objets connectés, l'heure des choix. https://www.frstrategie.org/sites/default/files/documents/publications/notes/2018/201815.pdf, 2018.

[12] Oscar Williams-Grut. Hackers once stole a casino's high-roller database through a thermometer in the lobby fish tank. https://www.businessinsider.com/hackers-stole-a-casinos-database-through-a-thermometer-in-the-lobby-fish-tank-2018-4?IR=T, 2018.

[13] Christian Cosquer and Julie Lanckriet. Les objets connectés et la défense. *Revue Defense Nationale*, (2) :97–103, 2016.

[14] Joe Thomas. How police unmasked "iceman" assassin behind one of britain's most notorious gangland murders. https://www.liverpoolecho.co.uk/news/liverpool-news/how-police-unmasked-iceman-assassin-15649613, 2019.

[15] Nicole Chavez. Arkansas judge drops murder charge in amazon echo case. https://edition.cnn.com/2017/11/30/us/amazon-echo-arkansas-murder-case-dismissed/index.html, 2017.

[16] Megan Cassidy. Fitbit offers key clue to slain san jose woman's alleged 90-year-old killer. https://www.sfchronicle.com/crime/article/Fitbit-offers-key-clue-to-slain-San-Jose-13266777.php, 2018.

[17] Urs Gasser, Nancy Gertner, Jack L Goldsmith, Susan Landau, Joseph S Nye, David O'Brien, Matthew G Olsen, Daphna Renan, Julian Sanchez, Bruce Schneider, et al. Don't panic : Making progress on the" going dark" debate. *Berkman Center Research Publication*, 2016.

[18] Le directeur du renseignement américain reconnaît s'intéresser aux objets connectés. https://www.lemonde.fr/pixels/article/2016/02/10/le-directeur-du-renseignement-americain-reconnait-s-interesser-aux-objets-connectes_4862587_4408996.html, 2016.

[19] François Bouchaud, Gilles Grimaud, and Thomas Vantroys. Iot forensic : identification and classification of evidence in criminal investigations. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, pages 1–9, 2018.

[20] François Bouchaud, Thomas Vantroys, Gilles Grimaud, and Pierrick Buret. Discovering connected objects in the criminal investigations. In *2020 International Conference on Omni-layer Intelligent Systems (COINS)*, pages 1–6. IEEE, 2020.

[21] François Bouchaud, Thomas Vantroys, and Gilles Grimaud. Evidence gathering in IoT criminal investigation. In *Digital Forensics and Cyber Crime*, pages 44–61. Springer, 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering.

[22] François Bouchaud, Gilles Grimaud, Thomas Vantroys, and Pierrick Buret. Digital investigation of iot devices in the criminal scene. *Journal of Universal Computer Science*, 25(9) :1199–1218, sep 2019. `http://www.jucs.org/jucs_25_9/digital_investigation_of_iot`.

[23] François Bouchaud. Bâtir une enquête 4.0. `https://fr.calameo.com/read/002719292 9f9a0d9dc123`, 2018.

[24] François Bouchaud. L'internet des objets à l'épreuve de la criminalistique numérique. `https://www.gendarmerie.interieur.gouv.fr/crgn/publications/revue-de-la-gendarmerie-nationale/revue-n-268`, 2020.

[25] Edmond Locard. *L'enquête criminelle et les méthodes scientifiques*. E. Flammarion, 1920.

[26] Pierre Margot. Traçologie : la trace, vecteur fondamental de la police scientifique. *Revue internationale de criminologie et de police technique et scientifique*, 67(1) :72–97, 2014.

[27] Paul L Kirk. The ontogeny of criminalistics. *J. Crim. L. Criminology & Police Sci.*, 54 :235, 1963.

[28] JD Nicol. The bachelor of sciences in criminalistics (program brochure), university of illinois, chicago, il, usa. 1972.

[29] Quon Yin Kwan. Inference of identity of source. *Ph. D. Dissertation. Department of Forensic Science., University of California*, 1976.

[30] Roger Cook, Ian W Evett, Graham Jackson, PJ Jones, and JA Lambert. A hierarchy of propositions : deciding which level to address in casework. *Science & Justice*, 4(38) :231–239, 1998.

[31] Roger Cook, Ian W Evett, Graham Jackson, PJ Jones, and JA Lambert. Case pre-assessment and review in a two-way transfer case. *Science & Justice*, 39(2) :103–111, 1999.

[32] Ian W Evett, G Jackson, and JA Lambert. More on the hierarchy of propositions : exploring the distinction between explanations and propositions. *Science & justice : journal of the Forensic Science Society*, 40(1) :3, 2000.

[33] Graham Jackson, Stella Jones, Gareth Booth, Christophe Champod, and Ian W Evett. The nature of forensic science opinion–a possible framework to guide thinking and prac-

tice in investigations and in court proceedings. *Science & justice : journal of the Forensic Science Society*, 46(1) :33–44, 2006.

[34] Charles Sanders Peirce. Collected papers, vol. 1–6, eds hartshorne and weiss. *Cambridge : Harvard University Press*, 1935 :7–8, 1931.

[35] Umberto Eco. The sign of three : Dupin, holmes, peirce advances in. 1983.

[36] Keith Inman and Norah Rudin. *Principles and practice of criminalistics : the profession of forensic science*. CRC Press, 2000.

[37] Luuc Van Der Horst, Kim-Kwang Raymond Choo, and Nhien-An Le-Khac. Process memory investigation of the bitcoin clients electrum and bitcoin core. *IEEE Access*, 5 :22385–22398, 2017.

[38] Xiaolu Zhang, Timothy T Yuen, and Kim-Kwang Raymond Choo. Experiential learning in digital forensics. In *Digital Forensic Education*, pages 1–9. Springer, 2020.

[39] Xiaodong Lin. File carving. In *Introductory Computer Forensics*, pages 211–233. Springer, 2018.

[40] Kai Shi, Ming Xu, Haoxia Jin, Tong Qiao, Xue Yang, Ning Zheng, Jian Xu, and Kim-Kwang Raymond Choo. A novel file carving algorithm for national marine electronics association (nmea) logs in gps forensics. *Digital Investigation*, 23 :11–21, 2017.

[41] Darren Quick and Kim-Kwang Raymond Choo. Iot device forensics and data reduction. *IEEE Access*, 6 :47566–47574, 2018.

[42] Christopher M Rondeau, Michael A Temple, and Juan Lopez. Industrial iot cross-layer forensic investigation. *Wiley Interdisciplinary Reviews : Forensic Science*, 1(1) :e1322, 2019.

[43] Konstantia Barmpatsalou, Tiago Cruz, Edmundo Monteiro, and Paulo Simoes. Current and future trends in mobile device forensics : A survey. *ACM Computing Surveys (CSUR)*, 51(3) :1–31, 2018.

[44] Ming Di Leom, Kim-Kwang Raymond Choo, and Ray Hunt. Remote wiping and secure deletion on mobile devices : A review. *Journal of forensic sciences*, 61(6) :1473–1492, 2016.

[45] Darren Quick and Kim-Kwang Raymond Choo. Pervasive social networking forensics : Intelligence and evidence from mobile device extracts. *Journal of Network and Computer Applications*, 86 :24–33, 2017.

[46] Xiaolu Zhang, Frank Breitinger, and Ibrahim Baggili. Rapid android parser for investigating dex files (rapid). *Digital Investigation*, 17 :28–39, 2016.

[47] Xiaolu Zhang, Ibrahim Baggili, and Frank Breitinger. Breaking into the vault : Privacy, security and forensic analysis of android vault applications. *Computers & Security*, 70 :516–531, 2017.

[48] MA Manazir Ahsan, Ainuddin Wahid Abdul Wahab, Mohd Yamani Idna Idris, Suleman Khan, Eric Bachura, and Kim-Kwang Raymond Choo. Class : Cloud log assuring soundness and secrecy scheme for cloud forensics. *IEEE Transactions on Sustainable Computing*, 2018.

[49] Suleman Khan, Abdullah Gani, Ainuddin Wahid Abdul Wahab, Mustapha Aminu Bagiwa, Muhammad Shiraz, Samee U Khan, Rajkumar Buyya, and Albert Y Zomaya. Cloud log forensics : Foundations, state of the art, and future directions. *ACM Computing Surveys (CSUR)*, 49(1) :1–42, 2016.

[50] Xiaolu Zhang, Justin Grannis, Ibrahim Baggili, and Nicole Lang Beebe. Frameup : an incriminatory attack on storj : a peer to peer blockchain enabled distributed storage system. *Digital Investigation*, 29 :28–42, 2019.

[51] Christian J D'Orazio and Kim-Kwang Raymond Choo. Circumventing ios security mechanisms for apt forensic investigations : A security taxonomy for cloud apps. *Future Generation Computer Systems*, 79 :247–261, 2018.

[52] Matthew C Stamm and KJ Ray Liu. Anti-forensics of digital image compression. *IEEE Transactions on Information Forensics and Security*, 6(3) :1050–1065, 2011.

[53] Christian DOrazio, Aswami Ariffin, and Kim-Kwang Raymond Choo. ios anti-forensics : How can we securely conceal, delete and insert data ? In *2014 47th Hawaii International Conference on System Sciences*, pages 4838–4847. IEEE, 2014.

[54] Brett Eterovic-Soric, Kim-Kwang Raymond Choo, Sameera Mubarak, and Helen Ashman. Windows 7 antiforensics : a review and a novel approach. *Journal of forensic sciences*, 62(4) :1054–1070, 2017.

[55] Richard Adams. *The advanced data acquisition model (ADAM) : a process model for digital forensic practice*. PhD thesis, Murdoch University, 2012.

[56] Electronic Crime Scene Investigation. A guide for first responders. *US Department of Justice, NCJ*, 187736, 2001.

[57] Eoghan Casey. *Digital evidence and computer crime : Forensic science, computers, and the internet*. Academic press, 2011.

[58] Maarten Van Horenbeeck. Technology crime investigation. *Archived from the original on*, 17, 2008.

[59] Jeffrey Voas. Networks of 'things'. *NIST Special Publication*, 800(183) :800–183, 2016.

[60] Christopher Greer, Martin Burns, David Wollman, and Edward Griffor. Cyber-physical systems and internet of things, 2019.

[61] JORF n°0008 du 11 janvier 2018 Texte n°135. Vocabulaire des télécommunications (liste de termes, expressions et définitions adoptés. `https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000036449955`.

[62] DHS. Us department of homeland security : Strategic principles for securing the internet of things (iot). 2016.

[63] Rec. UIT-T Y.2060 (06/2012). Présentation générale de l'internet des objets. `https://www.itu.int/rec/T-REC-Y.2060-201206-I/`.

[64] Commission des affaires économiques de l'Assemblée Nationale française n°4362. Rapport d'information déposé en application de l'article 145 du règlement par la commission des affaires économiques sur les objets connectés. `http://www.assemblee-nationale.fr/14/rap-info/i4362.asp`.

[65] Bruno Dorsemaine, Jean-Philippe Gaulier, Jean-Philippe Wary, Nizar Kheir, and Pascal Urien. Internet of things : a definition & taxonomy. In *2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies*, pages 72–77. IEEE, 2015.

[66] Carsten Maple. Security and privacy in the internet of things. *Journal of Cyber Policy*, 2(2) :155–184, 2017.

[67] Robert Hegarty, David J Lamb, Andrew Attwood, et al. Digital evidence challenges in the internet of things. In *INC*, pages 163–172, 2014.

[68] Nurul Huda Nik Zulkipli, Ahmed Alenezi, and Gary B Wills. Iot forensic : bridging the challenges in digital forensic and the internet of things. In *International Conference on Internet of Things, Big Data and Security*, volume 2, pages 315–324. SCITEPRESS, 2017.

[69] Francesco Servida and Eoghan Casey. Iot forensic challenges and opportunities for digital traces. *Digital Investigation*, 28 :S22–S29, 2019.

[70] Edewede Oriwoh, David Jazani, Gregory Epiphaniou, and Paul Sant. Internet of things forensics : Challenges and approaches. In *9th IEEE International Conference on Collaborative computing : networking, Applications and Worksharing*, pages 608–615. IEEE, 2013.

[71] Victor R Kebande and Indrakshi Ray. A generic digital forensic investigation framework for internet of things (iot). In *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, pages 356–362. IEEE, 2016.

[72] Bogdan Copos, Karl Levitt, Matt Bishop, and Jeff Rowe. Is anybody home ? inferring activity from smart home network traffic. In *2016 IEEE Security and Privacy Workshops (SPW)*, pages 245–251. IEEE, 2016.

[73] Shams Zawoad and Ragib Hasan. Faiot : Towards building a forensics aware eco system for the internet of things. In *2015 IEEE International Conference on Services Computing*, pages 279–284. IEEE, 2015.

[74] KM Sabidur Rahman, Matt Bishop, and Albert Holt. Internet of things mobility forensics. *de Proceedings of the 2016 Information Security Research and Education (INSuRE)*, 2016.

[75] Keyun Ruan, Joe Carthy, Tahar Kechadi, and Mark Crosbie. Cloud forensics. In *IFIP International Conference on Digital Forensics*, pages 35–46. Springer, 2011.

[76] Kim-Kwang Raymond Choo et al. Cloud computing : Challenges and future directions. *Trends and Issues in Crime and Criminal justice*, (400) :1, 2010.

[77] Diane Barrett and Greg Kipper. *Virtualization and forensics : A digital forensic investigator's guide to virtual environments.* Syngress, 2010.

[78] Christian Esposito, Aniello Castiglione, Florin Pop, and Kim-Kwang Raymond Choo. Challenges of connecting edge and cloud computing : A security and forensic perspective. *IEEE Cloud Computing*, 4(2) :13–17, 2017.

[79] Valentina Casola, Aniello Castiglione, Kim-Kwang Raymond Choo, and Christian Esposito. Healthcare-related data in the cloud : challenges and opportunities. *IEEE cloud computing*, 3(6) :10–14, 2016.

[80] Faheem Zafar, Abid Khan, Saif Ur Rehman Malik, Mansoor Ahmed, Adeel Anjum, Majid Iqbal Khan, Nadeem Javed, Masoom Alam, and Fuzel Jamil. A survey of cloud computing data integrity schemes : Design challenges, taxonomy and future trends. *Computers & Security*, 65 :29–49, 2017.

[81] M Edington Alex and R Kishore. Forensics framework for cloud computing. *Computers & Electrical Engineering*, 60 :193–205, 2017.

[82] Josiah Dykstra and Alan T Sherman. Understanding issues in cloud forensics : two hypothetical case studies. *UMBC Computer Science and Electrical Engineering Department*, 2011.

[83] Stephen Biggs and Stilianos Vidalis. Cloud computing : The impact on digital forensic investigations. In *2009 International Conference for Internet Technology and Secured Transactions,(ICITST)*, pages 1–6. IEEE, 2009.

[84] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, and Jin Li. Enabling public auditability and data dynamics for storage security in cloud computing. *IEEE transactions on parallel and distributed systems*, 22(5) :847–859, 2010.

[85] Yannan Li, Yong Yu, Geyong Min, Willy Susilo, Jianbing Ni, and Kim-Kwang Raymond Choo. Fuzzy identity-based data integrity auditing for reliable cloud storage systems. *IEEE Transactions on Dependable and Secure Computing*, 16(1) :72–83, 2017.

[86] Quratulain Alam, Saif UR Malik, Adnan Akhunzada, Kim-Kwang Raymond Choo, Saher Tabbasum, and Masoom Alam. A cross tenant access control (ctac) model for cloud computing : Formal specification and verification. *IEEE Transactions on Information Forensics and Security*, 12(6) :1259–1268, 2016.

[87] Vassil Roussev and Shane McCulley. Forensic analysis of cloud-native artifacts. *Digital Investigation*, 16 :S104–S113, 2016.

[88] Mark Taylor, John Haggerty, David Gresty, and Robert Hegarty. Digital evidence in cloud computing systems. *Computer law & security review*, 26(3) :304–308, 2010.

[89] Rajan Udeshi. Why you need forensics in an iot world. `https://www.forensicfocus.com/news/why-you-need-forensics-in-an-iot-world/`, 2017.

[90] Bhanu Prakash Kondapally. What is iot forensics and how is it different from digital forensics ? `https://cover2investigations.com/what-is-iot-forensics-and-how-is-it-different-from-digital-forensics/`, 2018.

[91] RC Joshi and Emmanuel S Pilli. *Fundamentals of Network Forensics*. Springer, 2016.

[92] Karen Kent, Suzanne Chevalier, Tim Grance, and Hung Dang. Guide to integrating forensic techniques into incident response. *NIST Special Publication*, 10(14) :800–86, 2006.

[93] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot) : A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7) :1645–1660, 2013.

[94] Isaac Amundson and Xenofon D Koutsoukos. A survey on localization for mobile wireless sensor networks. In *International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments*, pages 235–254. Springer, 2009.

[95] M Cypriani. Géopositionnement wi-fi autocalibré en milieu hétérogène. *Université de Franche-Comté*, 2012.

[96] Theodore S Rappaport et al. *Wireless communications : principles and practice*, volume 2. prentice hall PTR New Jersey, 1996.

[97] Vinita Daiya, Jemimah Ebenezer, SAV Satya Murty, and Baldev Raj. Experimental analysis of rssi for distance and position estimation. In *2011 International Conference on Recent trends in information technology (ICRTIT)*, pages 1093–1098. IEEE, 2011.

[98] Eduardo Cassano, Francesco Florio, Floriano De Rango, and Salvatore Marano. A performance comparison between roc-rssi and trilateration localization techniques for wpan sensor networks in a real outdoor testbed. In *2009 Wireless Telecommunications Symposium*, pages 1–8. IEEE, 2009.

[99] Frédéric Evennou. *Techniques et technologies de localisation avancées pour terminaux mobiles dans les environnements indoor*. PhD thesis, 2007.

[100] Youngjune Gwon, Ravi Jain, and Toshiro Kawahara. Robust indoor location estimation of stationary and mobile users. In *IEEE INFOCOM 2004*, volume 2, pages 1032–1043. IEEE, 2004.

[101] Paramvir Bahl and Venkata N Padmanabhan. Radar : An in-building rf-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064)*, volume 2, pages 775–784. Ieee, 2000.

[102] Fangfang Dong, Yiqiang Chen, Junfa Liu, Qiong Ning, and Songmei Piao. A calibration-free localization solution for handling signal strength variance. In *International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments*, pages 79–90. Springer, 2009.

[103] Paolo Barsocchi, Francesco Furfari, Paolo Nepa, and Francesco Potortì. Rssi localisation with sensors placed on the user. In *2010 International Conference on Indoor Positioning and Indoor Navigation*, pages 1–6. IEEE, 2010.

[104] Frédéric Lassabe, Philippe Canalda, and Pascal Chatonnay. Geolocalisation wifi et modeles de prediction de la mobilite dans les reseaux multimedia, 2009.

[105] Marco Altini, Davide Brunelli, Elisabetta Farella, and Luca Benini. Bluetooth indoor localization with multiple neural networks. In *IEEE 5th International Symposium on Wireless Pervasive Computing 2010*, pages 295–300. IEEE, 2010.

[106] Fabio Forno, Giovanni Malnati, and Giuseppe Portelli. Design and implementation of a bluetooth ad hoc network for indoor positioning. *IEE proceedings-Software*, 152(5) :223–228, 2005.

[107] Jan Blumenthal, Ralf Grossmann, Frank Golatowski, and Dirk Timmermann. Weighted centroid localization in zigbee-based sensor networks. In *2007 IEEE international symposium on intelligent signal processing*, pages 1–6. IEEE, 2007.

[108] Anne Ferréol. *Radio-goniométrie : modélisation, algorithmes, performances.* PhD thesis, 2005.

[109] Dominik Lieckfeldt, Jiaxi You, and Dirk Timmermann. Characterizing the influence of human presence on bistatic passive rfid-system. In *2009 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, pages 338–343. IEEE, 2009.

[110] Murat Torlak. Path loss. *New York : UTD*, 2002.

[111] Sinh Cong Lam and Kumbesan Sandrasegaran. Analytical and simulation performance of a typical user in random cellular network. *arXiv preprint arXiv :1607.03280*, 2016.

[112] Muhammad Sharjeel Zareen, Adeela Waqar, and Baber Aslam. Digital forensics : Latest challenges and response. In *2013 2nd National Conference on Information Assurance (NCIA)*, pages 21–29. IEEE, 2013.

[113] Oliver Hahm, Emmanuel Baccelli, Hauke Petersen, and Nicolas Tsiftes. Operating systems for low-end devices in the internet of things : a survey. *IEEE Internet of Things Journal*, 3(5) :720–734, 2015.

[114] Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, and Imrich Chlamtac. Internet of things : Vision, applications and research challenges. *Ad hoc networks*, 10(7) :1497–1516, 2012.

[115] Yongrui Qin, Quan Z Sheng, Nickolas JG Falkner, Schahram Dustdar, Hua Wang, and Athanasios V Vasilakos. When things matter : A survey on data-centric internet of things. *Journal of Network and Computer Applications*, 64 :137–153, 2016.

[116] Ameer Pichan, Mihai Lazarescu, and Sie Teng Soh. Cloud forensics : Technical challenges, solutions and comparative analysis. *Digital investigation*, 13 :38–57, 2015.

[117] Brian Carrier, Eugene H Spafford, et al. Getting physical with the digital investigation process. *International Journal of digital evidence*, 2(2) :1–20, 2003.

[118] AMC Gallop. Private practice, public duty. *Science & Justice*, 2(40) :104–108, 2000.

[119] Don Dovaston. The police perspective. *Science & Justice*, 2(40) :150–151, 2000.

[120] F Crispino. Computerized forensic assistance software (fas 1.0) for training and standardized investigation in distributed and disconnected services. *Forensic science international*, 132(2) :125–129, 2003.

[121] Conseil de l'Europe. Preuves électroniques dans les procédures civiles et administratives. `https://rm.coe.int/lignes-directrices-sur-les-preuves-electroniques-et-expose-des-motifs/1680968ab6`, 2019.

[122] Vrizlynn LL Thing, Kian-Yong Ng, and Ee-Chien Chang. Live memory forensics of mobile phones. *digital investigation*, 7 :S74–S82, 2010.

[123] Frank Adelstein. Live forensics : diagnosing your system without killing it first. *Communications of the ACM*, 49(2) :63–66, 2006.

[124] Stefan Vömel and Felix C Freiling. A survey of main memory acquisition and analysis techniques for the windows operating system. *Digital Investigation*, 8(1) :3–22, 2011.

[125] Hajime Inoue, Frank Adelstein, and Robert A Joyce. Visualization in testing a volatile memory forensic tool. *Digital Investigation*, 8 :S42–S51, 2011.

[126] Andrew Attwood, Madjid Merabti, Paul Fergus, and Omar Abuelmaatti. Sccir : Smart cities critical infrastructure response framework. In *2011 Developments in E-systems Engineering*, pages 460–464. IEEE, 2011.

[127] Cedric Chauvenet, Gerard Etheve, Mohamed Sedjai, and Manu Sharma. G3-plc based iot sensor networks for smartgrid. In *2017 IEEE International Symposium on Power Line Communications and its Applications (ISPLC)*, pages 1–6. IEEE, 2017.

[128] Jesse Kornblum. Preservation of fragile digital evidence by first responders. In *Digital Forensics Research Workshop (DFRWS)*, pages 1–11, 2002.

[129] Fernando Molina Granja and Glen D Rodríguez Rafael. The preservation of digital evidence and its admissibility in the court. *International Journal of Electronic Security and Digital Forensics*, 9(1) :1–18, 2017.

[130] Dominique Brezinski and Tom Killalea. Guidelines for evidence collection and archiving. *RFC3227, February*, 2002.

[131] Sarah Edwards and Heather Mahalik. Times a' ticking... to forensicate the apple watch ! `https://www.coursehero.com/file/36709855/Apple-Watch-Times-a-Tickin/`, 2015.

[132] Vladimir Katalov and Mattia Epifani. Apple watch forensics : Is it ever possible, and what is the profit ? `https://www.forensicfocus.com/news/apple-watch-forensics-is-it-ever-possible-and-what-is-the-profit-2/`, 2019.

[133] Dhenuka H Kasukurti and Suchitra Patil. Wearable device forensic : Probable case studies and proposed methodology. In *International Symposium on Security in Computing and Communication*, pages 290–300. Springer, 2018.

[134] Ibrahim Baggili, Jeff Oduro, Kyle Anthony, Frank Breitinger, and Glenn McGee. Watch what you wear : preliminary forensic analysis of smart watches. In *2015 10th International Conference on Availability, Reliability and Security*, pages 303–311. IEEE, 2015.

[135] Serim Kang, Soram Kim, and Jongsung Kim. Forensic analysis for iot fitness trackers and its application. *Peer-to-Peer Networking and Applications*, 13(2) :564–573, 2020.

[136] Courtney Grimes. Application analysis : Fitbit. `https://www.champlain.edu/Documents/LCDI/Application_Analysis__Fitbit.pdf`, 2017.

[137] Áine MacDermott, Stephen Lea, Farkhund Iqbal, Ibrahim Idowu, and Babar Shah. Forensic analysis of wearable devices : Fitbit, garmin and hetp watches. In *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–6. IEEE, 2019.

[138] Hyunji Chung, Jungheum Park, and Sangjin Lee. Digital forensic approaches for amazon alexa ecosystem. *Digital Investigation*, 22 :S15–S25, 2017.

[139] Shancang Li, Kim-Kwang Raymond Choo, Qindong Sun, William J Buchanan, and Jiuxin Cao. Iot forensics : Amazon echo as a use case. *IEEE Internet of Things Journal*, 6(4) :6487–6497, 2019.

[140] Ike Clinton, Lance Cook, and Shankar Banik. A survey of various methods for analyzing the amazon echo. *The Citadel, The Military College of South Carolina*, 2016.

[141] Jessica Hyde and Brian Moran. Alexa, are you skynet. *SANS Digital Forensics and Incident Response Summit*, 2017.

[142] Alex Akinbi and Thomas Berry. Forensic investigation of google assistant. *SN Computer Science*, 1(5) :1–10, 2020.

[143] Sean Tristan, Shally Sharma, and Robert Gonzalez. Alexa/google home forensics. In *Digital Forensic Education*, pages 101–121. Springer, 2020.

[144] Peter van Bolhuis and Cedric Van Bockhaven. Forensic analysis of chromecast and miracast devices. *Cybercrime and Forensics Project, Master's Program in System and Network Engineering, University of Amsterdam, Amsterdam, The Netherlands*, 2014.

[145] Nitesh K Bharadwaj and Upasna Singh. Acquisition and analysis of forensic artifacts from raspberry pi an internet of things prototype platform. In *Recent Findings in Intelligent Computing Techniques*, pages 311–322. Springer, 2019.

[146] Tanveer Zia, Peng Liu, and Weili Han. Application-specific digital forensics investigative model in internet of things (iot). In *Proceedings of the 12th International Conference on Availability, Reliability and Security*, pages 1–7, 2017.

[147] Jonathan T Rajewski. Internet of things forensics. *A Presentation at Enfuse*, 2016.

[148] Jacques Boucher and Nhien-An Le-Khac. Forensic framework to identify local vs synced artefacts. *Digital Investigation*, 24 :S68–S75, 2018.

[149] Arnoud Goudbeek, Kim-Kwang Raymond Choo, and Nhien-An Le-Khac. A forensic investigation framework for smart home environment. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 1446–1451. IEEE, 2018.

[150] Xiaolu Zhang, Kim-Kwang Raymond Choo, and Nicole Lang Beebe. How do i share my iot forensic experience with the broader community ? an automated knowledge sharing iot forensic platform. *IEEE Internet of Things Journal*, 6(4) :6850–6861, 2019.

[151] Yousef Amar, Hamed Haddadi, Richard Mortier, Anthony Brown, James Colley, and Andy Crabtree. An analysis of home iot network traffic and behaviour. *arXiv preprint arXiv :1803.05368*, 2018.

[152] Roman Ferrando and Paul Stacey. Classification of device behaviour in internet of things infrastructures : towards distinguishing the abnormal from security threats. In *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*, pages 1–7, 2017.

[153] Matias RP Santos, Rossana MC Andrade, Danielo G Gomes, and Arthur C Callado. An efficient approach for device identification and traffic classification in iot ecosystems. In *2018 IEEE Symposium on Computers and Communications (ISCC)*, pages 00304–00309. IEEE, 2018.

[154] Jingsha He, Chengyue Chang, Peng He, and Muhammad Salman Pathan. Network forensics method based on evidence graph and vulnerability reasoning. *Future Internet*, 8(4) :54, 2016.

[155] Patrick Neise. *Graph-based Event Correlation for Network Security Defense*. PhD thesis, The George Washington University, 2018.

[156] Chanyang Shin, Prerit Chandok, Ran Liu, Seth James Nielson, and Timothy R Leschke. Potential forensic analysis of iot data : an overview of the state-of-the-art and future possibilities. In *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 705–710. IEEE, 2017.

[157] Richard P Ayers, Sam Brothers, and Wayne Jansen. Guidelines on mobile device forensics. Technical report, 2014.

[158] S Brothers. How cell phone" forensic" tools actually work-proposed leveling system. In *Mobile Forensics World Conference, Chicago, Illinois*, 2009.

[159] Stephen Smalley and Trust Mechanisms R2X. The case for se android. *Linux Security Summit*, 2011.

[160] Hadeel Tariq Al-Rayes. Studying main differences between android & linux operating systems. *International Journal of Electrical & Computer Sciences IJECS-IJENS*, 12(05), 2012.

[161] Joseph T Sylve. Android memory capture and applications for security and privacy. 2011.

[162] Keonwoo Kim, Dowon Hong, Kyoil Chung, and Jae-Cheol Ryou. Data acquisition from cell phone using logical approach. *Proceedings of the world academy of science, engineering and technology*, 26, 2007.

[163] Ben Martini and Kim-Kwang Raymond Choo. An integrated conceptual digital forensic framework for cloud computing. *Digital Investigation*, 9(2) :71–80, 2012.

[164] Bharat Manral, Gaurav Somani, Kim-Kwang Raymond Choo, Mauro Conti, and Manoj Singh Gaur. A systematic survey on cloud forensics challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)*, 52(6) :1–38, 2019.

[165] Draft NISTIR. 8006 (2014) nist cloud computing forensic science challenges accessed at http ://csrc. nist. gov/publications/drafts/nistir-8006/draft_nistir_8006. pdf gary palmer (2001),"a road map for digital forensic research". Technical report, Technical Report DTR-T001-01, DFRWS, Report From the.

[166] James Cheney, Stephen Chong, Nate Foster, Seltzer Margo, and Stijn Vansummeren. Provenance : a future history. In *Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systema languages and applications*, pages 957–964, 2009.

[167] Yizhen Jia, Yinhao Xiao, Jiguo Yu, Xiuzhen Cheng, Zhenkai Liang, and Zhiguo Wan. A novel graph-based mechanism for identifying traffic vulnerabilities in smart home iot. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 1493–1501. IEEE, 2018.

[168] Henry P Romero, Kate A Remley, Dylan F Williams, and Chih-Ming Wang. Electromagnetic measurements for counterfeit detection of radio frequency identification cards. *IEEE Transactions on Microwave Theory and Techniques*, 57(5) :1383–1387, 2009.

[169] Smail Tedjini, Etienne Perret, Arnaud Vena, and Darine Kaddour. Mastering the electromagnetic signature of chipless rfid tags. In *Chipless and conventional radio frequency identification : Systems for ubiquitous tagging*, pages 146–174. IGI Global, 2012.

[170] Vladimir Brik, Suman Banerjee, Marco Gruteser, and Sangho Oh. Wireless device identification with radiometric signatures. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*, pages 116–127, 2008.

[171] Irwin O Kennedy and Alexandr M Kuzminskiy. Rf fingerprint detection in a wireless multipath channel. In *2010 7th International Symposium on Wireless Communication Systems*, pages 820–823. IEEE, 2010.

[172] Charles G Wheeler and Donald R Reising. Assessment of the impact of cfo on rf-dna fingerprint classification performance. In *2017 International Conference on Computing, Networking and Communications (ICNC)*, pages 110–114. IEEE, 2017.

[173] Saeed Ur Rehman, Kevin W Sowerby, Shafiq Alam, and Iman Ardekani. Radio frequency fingerprinting and its challenges. In *2014 IEEE Conference on Communications and Network Security*, pages 496–497. IEEE, 2014.

[174] Qiang Xu, Rong Zheng, Walid Saad, and Zhu Han. Device fingerprinting in wireless networks : Challenges and opportunities. *IEEE Communications Surveys & Tutorials*, 18(1) :94–104, 2015.

[175] Hua Fu, Samith Abeywickrama, Lihao Zhang, and Chau Yuen. Low-complexity portable passive drone surveillance via sdr-based signal processing. *IEEE Communications Magazine*, 56(4) :112–118, 2018.

[176] Adam C Polak and Dennis L Goeckel. Identification of wireless devices of users who actively fake their rf fingerprints with artificial data distortion. *IEEE Transactions on Wireless Communications*, 14(11) :5889–5899, 2015.

[177] Randall W Klein, Michael A Temple, and Michael J Mendenhall. Application of wavelet-based rf fingerprinting to enhance wireless network security. *Journal of Communications and Networks*, 11(6) :544–555, 2009.

[178] Ronald A Riley, James T Graham, Ryan M Fuller, Rusty O Baldwin, and Ashwin Sampathkumar. Extraction and validation of algorithms based on analog side-channels. In *Cyber Sensing 2017*, volume 10185, page 1018506. International Society for Optics and Photonics, 2017.

[179] Devin W Spatz, Devin A Smarra, and Igor V Ternovskiy. Preliminary classification results of rf emission based feature extraction in internet of things devices. In *Cyber Sensing 2018*, volume 10630, page 106300E. International Society for Optics and Photonics, 2018.

[180] Sandra Siby, Rajib Ranjan Maiti, and Nils Tippenhauer. Iotscanner : Detecting and classifying privacy threats in iot neighborhoods. *arXiv preprint arXiv :1701.05007*, 2017.

[181] Bruhadeshwar Bezawada, Maalvika Bachani, Jordan Peterson, Hossein Shirazi, Indrakshi Ray, and Indrajit Ray. Iotsense : Behavioral fingerprinting of iot devices. *arXiv preprint arXiv :1804.03852*, 2018.

[182] N Aluthge. Iot device fingerprinting with sequence-based features. *Master's thesis*, 2017.

[183] Azeez J Bhavnagarwala, Xinghai Tang, and James D Meindl. The impact of intrinsic device fluctuations on cmos sram cell stability. *IEEE journal of Solid-state circuits*, 36(4) :658–665, 2001.

[184] Eric Adler, John K DeBrosse, Stephen F Geissler, Steven J Holmes, Mark D Jaffe, Jeffrey B Johnson, CW Koburger, Jerome B Lasky, Brian Lloyd, Glen L Miles, et al. The evolution of ibm cmos dram technology. *IBM Journal of Research and Development*, 39(1.2) :167–188, 1995.

[185] Paul Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis, advances in cryptology-crypto'99. In *Proc. 19th Annual International Cryptology Conf*, pages 388–397, 1999.

[186] Jasper GJ van Woudenberg, Marc F Witteman, and Bram Bakker. Improving differential power analysis by elastic alignment. In *Cryptographers' Track at the RSA Conference*, pages 104–119. Springer, 2011.

[187] Nadir Cherifi, Thomas Vantroys, Alexandre Boe, Colombe Herault, and Gilles Grimaud. Automatic inference of energy models for peripheral components in embedded systems. In *2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)*, pages 120–127. IEEE, 2017.

[188] Jean-Jacques Quisquater and David Samyde. Electromagnetic analysis (ema) : Measures and counter-measures for smart cards. In *International Conference on Research in Smart Cards*, pages 200–210. Springer, 2001.

[189] Donald H Habing. The use of lasers to simulate radiation-induced transients in semiconductor devices and circuits. *IEEE Transactions on Nuclear Science*, 12(5) :91–100, 1965.

[190] Dan Boneh, Richard A DeMillo, and Richard J Lipton. On the importance of checking cryptographic protocols for faults. In *International conference on the theory and applications of cryptographic techniques*, pages 37–51. Springer, 1997.

[191] Sergei P Skorobogatov and Ross J Anderson. Optical fault induction attacks. In *International workshop on cryptographic hardware and embedded systems*, pages 2–12. Springer, 2002.

[192] Nicolas Moro, Amine Dehbaoui, Karine Heydemann, Bruno Robisson, and Emmanuelle Encrenaz. Electromagnetic fault injection : towards a fault model on a 32-bit microcontroller. In *2013 Workshop on Fault Diagnosis and Tolerance in Cryptography*, pages 77–88. IEEE, 2013.

[193] Amine Dehbaoui, Jean-Max Dutertre, Bruno Robisson, and Assia Tria. Electromagnetic transient faults injection on a hardware and a software implementations of aes. In *2012 Workshop on Fault Diagnosis and Tolerance in Cryptography*, pages 7–15. IEEE, 2012.

[194] Chong Hee Kim and Jean-Jacques Quisquater. New differential fault analysis on aes key schedule : Two faults are enough. In *International Conference on Smart Card Research and Advanced Applications*, pages 48–60. Springer, 2008.

[195] Nidhal Selmane, Sylvain Guilley, and Jean-Luc Danger. Practical setup time violation attacks on aes. In *2008 Seventh European Dependable Computing Conference*, pages 91–96. IEEE, 2008.

[196] Alessandro Barenghi, Luca Breveglieri, Israel Koren, and David Naccache. Fault injection attacks on cryptographic devices : Theory, practice, and countermeasures. *Proceedings of the IEEE*, 100(11) :3056–3076, 2012.

[197] Sebanjila K Bukasa, Ronan Lashermes, Jean-Louis Lanet, and Axel Leqay. Let's shock our iot's heart : Armv7-m under (fault) attacks. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, pages 1–6, 2018.

[198] Jean Arlat, Yves Crouzet, Johan Karlsson, Peter Folkesson, Emmerich Fuchs, and Günther H Leber. Comparison of physical and software-implemented fault injection techniques. *IEEE Transactions on Computers*, 52(9) :1115–1133, 2003.

[199] J Ross. Anderson. security engineering : A guide to building dependable distributed systems, 2008.

[200] David Brumley and Dan Boneh. Remote timing attacks are practical. *Computer Networks*, 48(5) :701–716, 2005.

[201] Robert Beverly. A robust classifier for passive tcp/ip fingerprinting. In *International Workshop on Passive and Active Network Measurement*, pages 158–167. Springer, 2004.

[202] Andrew M White, Austin R Matthews, Kevin Z Snow, and Fabian Monrose. Phonotactic reconstruction of encrypted voip conversations : Hookt on fon-iks. In *2011 IEEE Symposium on Security and Privacy*, pages 3–18. IEEE, 2011.

[203] Brandon Niemczyk and Prasad Rao. Identification over encrypted channels. *BlackHat USA*, 2014.

[204] Daniel Genkin, Adi Shamir, and Eran Tromer. Rsa key extraction via low-bandwidth acoustic cryptanalysis. In *Annual Cryptology Conference*, pages 444–461. Springer, 2014.

[205] Dmitri Asonov and Rakesh Agrawal. Keyboard acoustic emanations. In *IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004*, pages 3–11. IEEE, 2004.

[206] Li Zhuang, Feng Zhou, and J Doug Tygar. Keyboard acoustic emanations revisited. *ACM Transactions on Information and System Security (TISSEC)*, 13(1) :1–26, 2009.

[207] Michael Backes, Markus Dürmuth, Sebastian Gerling, Manfred Pinkal, and Caroline Sporleder. Acoustic side-channel attacks on printers. In *USENIX Security symposium*, pages 307–322, 2010.

[208] Adi Shamir and Eran Tromer. Acoustic cryptanalysis : On nosy people and noisy machines. eurocrypt 2004 rump session, 2004.

[209] Michael LeMay and Jack Tan. Acoustic surveillance of physically unmodified pcs. In *Security and Management*, pages 328–334, 2006.

[210] Daniel Genkin, Mihir Pattani, Roei Schuster, and Eran Tromer. Synesthesia : Detecting screen content via remote acoustic side channels. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 853–869. IEEE, 2019.

Résumé

Avec le développement des écosystèmes connectés à Internet, la recherche de données dans un environnement numérique par l'enquêteur judiciaire constitue une tâche de plus en plus ardue. Elle est un véritable défi en particulier par l'hétérogénéité des objets à étudier. À cette affirmation, il convient d'y ajouter l'absence de standardisation des architectures de communication et des remontées de données, des dépendances entre les dispositifs connectés et une dispersion de l'information. Dans cette thèse de doctorat, nous proposons d'adapter l'approche traditionnelle de l'investigation numérique aux contraintes de l'Internet des objets. Nous développons des méthodologies et des outils d'appréhension et d'analyse de l'environnement connecté pour les praticiens du judiciaire. Nous partons du principe que la scène de crime constitue un tout connecté et non un agrégat d'objets numériques. Elle contient des données clefs dans la compréhension et la contextualisation d'un événement ou d'un phénomène passé, éléments de preuve pour le procès pénal.

L'investigation numérique est une « *science appliquée pour identifier un incident, collecter, examiner et analyser des données tout en préservant l'intégrité de l'information et en maintenant une chaîne de contrôle stricte pour les données* » (National Institute of Standards and Technology). Face à une scène de crime, l'enquêteur cherche à comprendre l'événement criminel, en examinant les traces figées ou emprisonnées dans le support physique et/ou dans une partie déportée sur le Cloud. Nos travaux développent un processus d'identification rapide du phénomène selon quatre phases : détection, localisation, reconnaissance des objets et recoupement de l'information. Il est enrichi d'outils de recherche de traces radioélectriques : simple capteur et réseau maillé multi-capteur. Cette démarche est construite autour de la problématique de l'appréhension d'un environnement connecté multiforme, contenant des dispositifs pas toujours visibles ou identifiables lors d'une approche terrain. Nous intégrons dans notre étude la stratégie de la collecte des équipements. Le défi réside dans la capacité à extraire un ou plusieurs objets connectés, sans compromettre les données stockées, pour les placer dans un environnement contrôlé et sécurisé. L'objet est maintenu dans un état garantissant la non-altération ou la perte des données. L'étude regroupe une première phase de compréhension de l'environnement physique et des dépendances. Elle cherche à déterminer les mécanismes de migration de l'information vers les plates-formes en ligne et à isoler les groupes d'objets en déstructurant avec intelligence les connexions. Les dispositifs sont extraits, puis conditionnés et scellés au regard de leurs caractéristiques techniques et de l'infrastructure connectée. Puis, nous approfondissons l'exploitation de l'information collectée par des méthodes forensiques. La donnée est alors analysée selon les axes temporels, spatiaux et contextuels. Nous proposons par ailleurs une classification et une priorisation de la structure connectée en fonction des caractéristiques de la donnée recherchée. Les travaux donnent une lecture du cycle de vie de la donnée au sein de l'infrastructure de l'Internet des Objets.

Dans une approche prospective, nous approfondissons les questions de l'identification fine de l'objet connecté en fonction des caractéristiques du matériel et du logiciel. L'émission acoustique de l'électronique apparaît comme une propriété physique pertinente dans l'étude des équipements. Cet attribut complète notre palette d'outils dans l'identification des objets connectés.

Mots-clés: Informatique, Internet des objets, Investigation, Criminalistique numérique, Identification

Abstract

With the development of the Internet of Things, searching for data in a digital environment is an increasingly difficult task for the forensic investigator. It is a real challenge, especially given the heterogeneity of the connected objects. There is a lack of standardization in communication architectures and data management policies. It is accompanied by dependencies between connected ecosystems, especially through hidden links and fragmented information. In this thesis, we suggest adjusting the traditional approach of digital investigation to the constraints of the Internet of Things. We develop methodologies and tools to understand and analyze the connected environment. We assume that the crime scene is a connected whole and not an aggregate of independent digital objects. It contains key data for understanding and contextualizing a past event or phenomenon as evidence for the criminal trial. Digital forensics is considered to be the « *application of science to the identification, collection, examination, and analysis, of data while preserving the integrity of the information and maintaining a strict chain of custody for the data* » (National Institute of Standards and Technology). Faced with a crime scene, the investigator seeks to understand the criminal event. He examines the data stored in the physical medium and/or in a remote part of the cloud. Our work develops a process of rapid identification of the phenomenon according to four phases : detection, localization, object recognition and information crosschecking. It is enriched with radio signature search tools : single-sensor and multi-sensor mesh network. This approach is built around the problem of apprehending a multiform connected environment, containing devices that are not always visible or identifiable during a field approach. We integrate in our study the strategy of equipment collection. The challenge lies in the ability to extract one or more connected objects, without compromising the stored data, to place them in a controlled and secure environment. The object is maintained in a state that guarantees the non-alteration or loss of data. The study includes a first phase of understanding the physical environment and dependencies. It seeks to determine the mechanisms of information migration to online platforms and to isolate groups of objects by intelligently breaking the connections. Devices are extracted, then packaged and sealed according to their technical characteristics and the connected infrastructure. We then deepen the exploitation of the information collected using forensic methods. The data is then analyzed according to temporal, spatial and contextual axes. We also propose a classification and a prioritization of the connected structure according to the characteristics of the desired data. The work gives a reading of the life cycle of the data within the Internet of Things infrastructure. In a prospective approach, we deepen the questions of the fine identification of the connected object according to these hardware and software characteristics. The acoustic signature of electronics appears as a relevant physical property in the study of equipment. This feature completes our range of tools in the identification of connected objects.

Keywords: Computer science, Internet of Things, Investigation, Digital Forensic, Identification