**L'intégration dans le cycle du renseignement d'une discipline en évolution: le ROSO (renseignement d'origine sources ouvertes)**

*Par le Capitaine Guerric GROSJEAN, stagiaire EMST,*
*en formation cyberdéfense à Telecom Paris*

Les opérations militaires deviennent de plus en plus complexes avec la multiplication des paramètres entrant en jeu (géopolitiques, sociaux…) et l'inbrication des enjeux civils et militaires. Les informations disponibles en amont se densifient et augmentent souvent le «brouillard de la guerre». Dès lors, la préparation des opérations requiert le recueil et l'interprétation des informations pertinentes pour en extraire le renseignement utile en un temps toujours plus contraint. Cela, afin de garantir une prise de décision la plus efficace.

Le renseignement d'origine cybernétique (ROC) et plus précisément sa composante «sources ouvertes» (ROSO) est un ensemble de méthodes consistant à traiter ces informations, qui sont accessibles à tous car non protégées, de la manière la plus efficace possible. Avec l'avènement des technologies de l'information et de la communication, des informations cruciales circulent, quel que soit le théâtre (intérieur comme extérieur à nos frontières), par des voies électroniques (web, réseaux et médias sociaux…). Dès lors le ROSO devient essentiel aux opérations et nécessite la maitrise de techniques en évolution souvent plus rapide que le rythme des opérations elles-mêmes.

Le ROSO est une discipline dont l'évolution est concomitante à celle d'Internet. Ses différents aspects sont aujourd'hui progressivement intégrés dans le cycle du renseignement et s'inscrivent pleinement dans les actions au sein du champ d'action immatériel évoqué dans la vision stratégique.

## • Le ROSO, une discipline en évolution

Le ROSO n'est pas une discipline très récente mais elle s'est révélée récemment. En 2006, Amnesty International a rendu public un rapport sur les mouvements extra-judiciaires de prisonniers liés au terrorisme. Ce rapport est le fruit de recueils et d'analyse de documents en source ouverte tels que des données de vols internationaux d'avions civils. Cette divulgation à bien évidemment eu des répercussions sur l'opinion publique et l'impact a été non négligeable pour les autorités. C'est au cours de cette période qu'un *Open Source Center* a été créé par la CIA à partir du *Foreign Broadcast Information Service* (service qui recueille et traduit les informations diffusées sur les radios et journaux du monde entier).

Cet exemple montre effectivement que, au sein de cette discipline, l'information doit absolument provenir de médias accessibles à tous. L'information disponible sur le web est notamment en constante évolution dans la mesure où les fournisseurs professionnels d'informations y ont tous migré. Elle se joint aux multiples blogs, podcast ou autres supports de diffusion. Ainsi, nombres d'informations hétérogènes (personne privée, institution, …) se retrouvent consultables par n'importe qui et n'importe quand, ce qui permet notamment de les recouper.

Le ROSO consiste à récupérer des données, les regrouper, les ranger et les interpréter. Il peut être utilisé comme moyen de surveillance et d'alerte et donc peut servir à des fins bienveillantes. Cela semble le cas pour le groupe Bellingcat. Il s'agit d'un groupe de chercheurs, d'enquêteurs et de journalistes qui utilisent les sources ouvertes pour enquêter sur des crimes ou sur l'emploi de l'arme chimique dans le monde.

Cependant, ces techniques peuvent aussi être employées à des fins malveillantes Par exemple la revente de données personnelles sur Internet peut être à l'origine d'activités frauduleuses telles que l'hameçonnage ciblé (*spear phising*).

- **L'intégration du ROSO dans le cycle du renseignement**

  - ***La phase de recueil***

La première phase permet le recueil d'informations. Il est nécessaire d'identifier et de cribler l'ensemble des acteurs d'un théâtre. Au XXI$^{ème}$ siècle, ces acteurs sont non seulement les forces en présence mais également les entreprises, les ONG, les personnes privées qui peuvent avoir une influence sur les opérations. Or, ces entités ont sans aucun doute laissé des informations sur Internet, les réseaux sociaux ou d'autres supports électroniques. En cela, le ROSO permet d'effectuer un premier renseignement nécessaire à l'établissement du contexte de toute opération et constitue un appui à la décision. La notion de temps est alors dimensionnante pour la pertinence d'un tel procédé.

Effectivement, les outils actuels, toujours plus nombreux et performants, vont permettre de recueillir les informations de façon rapide et automatisée. Outre les moteurs de recherche, il existe des méthodes que les néophytes peuvent presque aisément mettre en application. Il est ainsi aisé d'utiliser ce que l'on appelle en informatique un «*scraper*». Il s'agit d'extraire les informations d'un ou plusieurs sites web de manière automatique par des scripts (petits programmes informatiques). Ces programmes peuvent être élaborés par une personne ayant des bases en programmation mais ils peuvent aussi être disponibles en code source ouverte (*open source*) sur Internet, c'est-à-dire téléchargeables en quelques secondes. Par exemple, «Instagram Scraper» va permettre d'extraire diverses informations d'un profil Instagram (amis, photos, …) sans avoir à se connecter au profil. L'utilisateur va seulement entrer le nom du profil et le recueil sera automatisé. Il récupèrera toutes les informations dans un fichier unique qu'il pourra exploiter manuellement ou de manière automatisée (analyse automatisée d'images).

D'une autre façon, les agrégateurs de flux RSS permettent de faire de la veille. Un flux RSS est une information contenue dans une page web. Lorsque le contenu est mis à jour, une nouvelle information (un nouveau flux) est envoyée. L'agrégateur va permettre de regrouper toutes les nouvelles informations diffusées. Il est aisé, via cette solution, de recevoir automatiquement, tous les articles qui traitent de telle information sur tel site de tel pays. Cette méthode permet, sans prendre trop de temps, de surveiller les médias d'un pays en particulier pour y suivre la situation géopolitique.

Enfin, l'outil Shodan, très complet, met en exergue les vulnérabilités liées au recueil d'informations. En effet, cet outil permet de scanner et recueillir des informations sur tout objet électronique communiquant. Il peut s'agir de technologies opérationnelles telles que centrales électriques ou usines manufacturières via leurs caméras, capteurs ou dispositifs de sécurité. Mais il permet aussi de recueillir des informations sur les milliards d'objets connectés. Ainsi, avec une licence de quelques dizaines de dollars il est possible de faire une infinité de requêtes (demander des informations) sur des montres connectées pour obtenir leur position. Cela remet en mémoire les conséquences de l'utilisation de l'application Strava – application mobile utilisée pour enregistrer des activités sportives via GPS – qui a permis notamment de dévoiler des bases américaines sur des théâtres d'opération extérieurs dans la mesure où toutes les informations étaient collectées et diffusées sur le web. Un utilisateur pouvait facilement les regrouper et identifier une densité importante de joggeurs en plein désert.

  - ***La phase d'analyse***

Après recueil et tri, l'analyse peut être effectuée manuellement. L'exemple de la vidéo «Anatomy of a killing» de la BBC est pertinent et montre une manière d'analyser un média, en l'occurrence une vidéo. Dans un premier temps, la zone de l'évènement a été identifiée en analysant les courbes du relief visibles sur la vidéo et en confondant ces courbes avec celles disponibles sur un logiciel de cartographie. Ensuite, ce sont les emplacements de bâtiments et

d'une route qui ont été reconnus grâce à des images satellites. Les images satellites étant datées, elles ont également permis, par une brève analyse de dater les évènements. L'analyse des ombres a, quant à elle, permis de réduire la plage mensuelle. Enfin, l'analyse d'un reportage antérieur a conduit à l'identification d'un avant-poste de l'armée camerounaise. L'analyse du son a ensuite aidé à l'identification des surnoms. Ces surnoms ont été analysés et comparés avec des profils de réseaux sociaux afin d'identifier les acteurs. Cette analyse a ainsi été le fruit de recoupements d'informations collectées par des moyens électroniques et analysés en partie par des opérateurs.

La notion de temps est cependant cruciale dans les opérations. Dès lors, des outils d'analyses fondés sur l'intelligence artificielle se développent et vont permettre de gagner un temps précieux. L'exemple de la startup américaine Clearview AI montre les possibilités offertes. Leur solution logicielle permet d'identifier des visages grâce à une base de données composée de milliards de clichés collectés sur Internet. Cette solution, controversée (atteinte aux libertés individuelles), est actuellement employée par les forces de l'ordre américaines et montre une redoutable efficacité. Considérant que nombres d'opérations sont filmées et diffusées sur Internet, cette solution permettrait d'identifier rapidement les acteurs dont la plupart auraient un profil sur un réseau social.

- ***La phase de diffusion***

La phase de diffusion repose aujourd'hui encore sur l'expertise de l'analyste. Cependant, des outils se développent également dans ce domaine pour contextualiser et enrichir des données qui deviendront de véritables renseignements. Le projet MISP montre l'organisation qui se met en place. En effet, il s'agit d'une plateforme d'échange d'utilitaires (programmes entre autres) et de documentation pour la *Cyber Threat Intelligence* (CTI). La CTI est une discipline fondée sur des techniques de renseignement visant à la collecte et l'organisation de toutes les informations liées aux menaces du cyberespace. Diverses solutions sont mises en place et améliorées collectivement en vue d'organiser et de diffuser de manière ordonnée des informations collectées, analysées et recoupées.

Certaines solutions sont de réels appuis dans l'organisation des données. Le logiciel Maltego est ainsi spécialisé dans la découverte des relations entre les personnes, les entreprises, les domaines et les informations publiques sur Internet (Il permet de présenter des quantités parfois énormes d'informations découvertes sous forme de tableaux et de graphiques faciles à lire. Sont alors dévoilées des relations cachées entre des personnes, des entreprises ou autres entités.

## • Perspectives

D'une manière générale, le volume des applications tend à croitre rapidement. Chaque type de renseignement possède une application permettant de le recueillir, de l'analyser et de le diffuser de façon efficace (voir schéma ci-contre). Cette discipline intéresse de plus en plus d'entités qui se regroupent afin de partager leurs savoir et donc d'augmenter l'efficacité du cycle de renseignement. S'il existe des groupes de passionnés – par exemple OSINT FR – bienveillants, il existe indéniablement des entités malveillantes capables d'employer divers outils informatiques dans le but de nuire aux opérations. Dès lors, il semble indispensable de développer ses propres outils et ses connaissances dans un domaine clé et en constante évolution.

La vision stratégique du chef d'état-major de l'armée de Terre montre explicitement que les actions dans le champ immatériel doivent s'inscrire dans une stratégie globale. Ainsi, une Task Force Cyber aura toute légitimité pour œuvrer dans le cyberespace. Sa composante

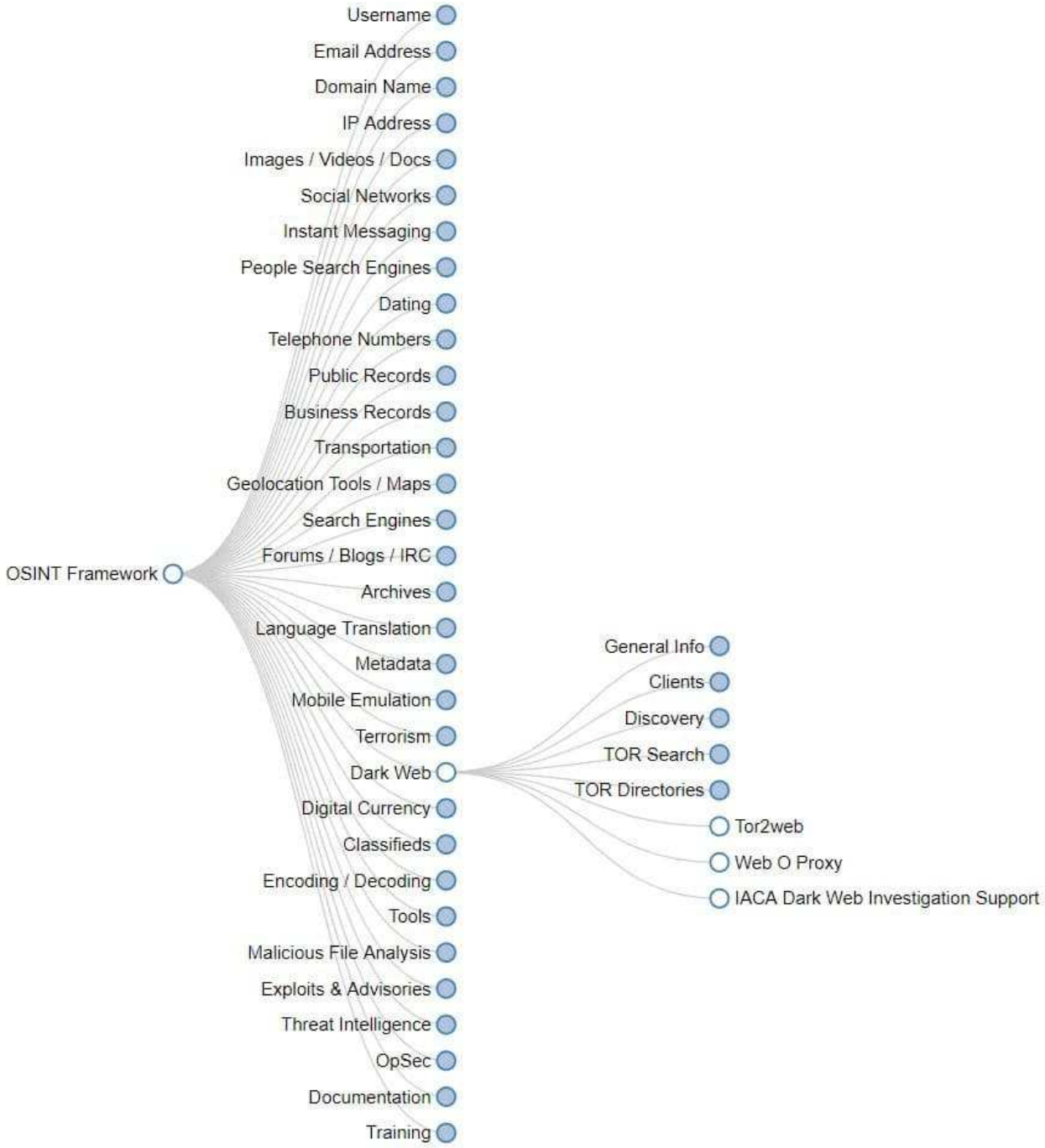
renseignement pourra s'inspirer de l'ensemble des techniques évoquées supra pour obtenir du renseignement ou effectuer des opérations d'influence en amont d'opérations hybrides.

## • Conclusion

Le renseignement d'origine sources ouvertes (ROSO) n'est pas une discipline nouvelle mais l'avènement de l'intelligence artificielle et la multitude d'algorithmes disponibles tendent à rendre cette discipline très efficace. La frontière entre ROSO et emploi malveillant de codes informatiques est poreuse. Des entités malveillantes possèdent ainsi une multitude de moyens pour récupérer des informations et les utiliser contre les forces armées ou les forces de sécurité intérieure. L'action dans le champ immatériel prend alors toute sa dimension et l'acquisition de la supériorité dans les domaines afférents tels que la CTI, le ROSO où la lutte informatique offensive est indispensable.

***Références :***

[1]https://cf2r.org/cyberrens/le-renseignement-via-les-sources-ouvertes-osint-une-nouvelle-discipline/
[2]https://fr.bellingcat.com/
[3]https://www.cairn.info/revue-defense-nationale-2017-5-page-161.htm?contenu=bibliographie
[4]https://clearview.ai/
[5]https://osintframework.com/


OSINT Framework
Username
Email Address
Domain Name
IP Address
Images / Videos / Docs
Social Networks
Instant Messaging
People Search Engines
Dating
Telephone Numbers
Public Records
Business Records
Transportation
Geolocation Tools / Maps
Search Engines
Forums / Blogs / IRC
Archives
Language Translation
Metadata
Mobile Emulation
Terrorism
Dark Web
Digital Currency
Classifieds
Encoding / Decoding
Tools
Malicious File Analysis
Exploits & Advisories
Threat Intelligence
OpSec
Documentation
Training
General Info
Clients
Discovery
TOR Search
TOR Directories
Tor2web
Web O Proxy
IACA Dark Web Investigation Support