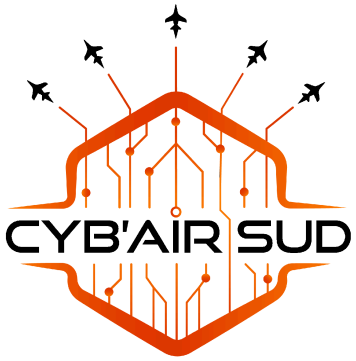


Behavioral Detection of Disk Activities of Ransomware and Synthetic Data Generation to Anticipate Future Attacks

CFP 2025 — English Version

Speaker: (to be confirmed)

Classification and dissemination

TLP: **GREEN French** or **English**.

Title

Behavioral Detection of Disk Activities of Ransomware and Synthetic Data Generation to Anticipate Future Attacks

Duration

30 minutes

Short Description

In a context where ransomware evolve faster than our defenses, this presentation presents our work on real-time detection of suspicious disk behaviors using 28 key indicators. We also explore how to generate synthetic data to simulate novel attacks, thereby strengthening proactive detection capabilities and reducing risk to critical systems.

Detailed Description

The Stakes: Anticipating Ransomware to Protect Critical Systems

Ransomware poses a major threat to digital infrastructures, able to encrypt data in minutes and cause massive disruptions. Traditional signature-based methods are reactive and often ineffective against emerging variants. Our objective is to develop proactive real-time detection based on behavioral analysis of disk activities to stop attacks before irreversible

damage occurs. While examples focus on ransomware, the approach is directly extensible to critical environment monitoring and C2 (Command and Control) detection. This talk highlights our work and our innovative approach to synthetic data generation, enabling anticipation of attack scenarios that do not yet exist.

Our Work: Intelligent Detection of Disk Activities for Early Detection

Our research focuses on discreet and efficient monitoring of disk activity, capturing program actions (reads, writes, file modifications) every 50 operations. This rate optimizes performance — avoiding system overload while reliably capturing suspicious signals independent of clock-based sampling, easing comparisons across environments.

We evaluate 28 indicators grouped into clear categories that form a behavioral fingerprint to identify anomalies related to ransomware:

— **Basic actions**: number of reads, writes, file opens, and volume of data handled;
— **Entropy**: measure of file "disorder", which spikes during malicious encryption;
— **File manipulation**: monitoring deletions, renames and suspicious modifications;
— **Targeted file types**: focus on sensitive extensions such as .doc, .pdf, .zip and executables;
— **Propagation**: analysis of dispersion across folders revealing typical ransomware clusters;
— **System information**: program identifier, number of processes and classification label (ransomware or not).

Our reference dataset includes 268,748 records of normal behavior, which serve as the baseline for detecting deviations.

Synthetic Data Generation: Preparing for the Unknown

To overcome the limitations of real data — incomplete, outdated and lacking variety — we generate realistic synthetic scenarios. Trained on authentic data, these models create hypothetical attacks that:

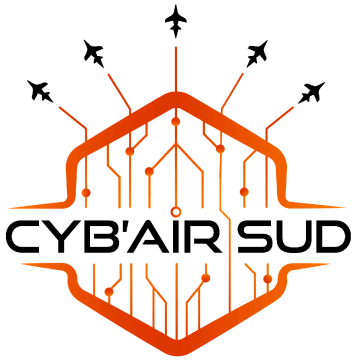
1. simulate previously unseen threats;
2. combine techniques in novel ways;
3. test the boundaries of detection.

We ensure realism by preserving temporal logic, natural relations between indicators (e.g., entropy increase during renames) and credible propagation patterns. This approach turns defense into a proactive strategy, increasing resilience against tomorrow's ransomware.

Other Information

— Language: French
— TLP: GREEN
— Recording / Redistribution: Allowed

— Duration: 30 minutes



Détection Comportementale des Activités sur Disque des Ransomwares et Génération de Données Synthétiques pour Anticiper les Attaques Futures

CFP 2025 — Version Française

Intervenant : (à confirmer)

Remarque concernant la participation

Il reste à confirmer notre participation physique (Salon étant éloigné de Pau et Paris). Pouvez-vous préciser si vous prenez en charge les déplacements ? À défaut, la participation en visioconférence est-elle possible ?

Classification et diffusion

TLP : **VERTE** — diffusion autorisée avec notre accord. La présentation peut être donnée en **français** ou en **anglais**.

Titre de la conférence

Détection Comportementale des Activités sur Disque des Ransomwares et Génération de Données Synthétiques pour Anticiper les Attaques Futures

Durée

30 minutes

Description concise

Dans un contexte où les ransomwares évoluent plus vite que nos défenses, cette présentation expose nos travaux sur la détection en temps réel des comportements suspects sur disque via 28 indicateurs clés. Nous explorons également la génération de données synthétiques pour simuler des attaques inédites, renforçant ainsi la détection proactive et minimisant les risques pour les systèmes critiques.

Résumé détaillé

Les enjeux : anticiper les ransomwares pour protéger les systèmes critiques

Les ransomwares représentent une menace majeure pour les infrastructures numériques, capables de chiffrer des données en quelques minutes et de provoquer des disruptions massives. Les méthodes traditionnelles basées sur les signatures sont réactives et souvent inefficaces face aux variantes émergentes. Notre objectif est de développer une détection proactive en temps réel, fondée sur l'analyse comportementale des activités sur disque, pour arrêter les attaques avant qu'elles ne causent des dommages irréversibles. Les exemples illustrent le cas des ransomwares, mais la méthode est directement extensible à la surveillance d'environnements critiques et à la détection de C2. Cette présentation met en lumière nos travaux et notre approche innovante de génération de données synthétiques permettant d'anticiper des scénarios d'attaques encore inexistants.

Nos travaux : une détection intelligente des activités sur disque pour une détection précoce

Nos recherches portent sur une surveillance discrète et efficace des activités sur disque, capturant les actions des programmes (lectures, écritures, modifications de fichiers) toutes les 50 opérations. Ce rythme optimise les performances : il évite de surcharger le système tout en capturant de manière fiable les signaux suspects, indépendamment de l'horloge, facilitant les comparaisons entre environnements.

Nous évaluons 28 indicateurs regroupés en catégories claires, formant une empreinte comportementale unique pour identifier les anomalies liées aux ransomwares :

- **Actions de base** : nombre de lectures, écritures, ouvertures de fichiers et volume de données manipulées ;
- **Entropie** : mesure du "désordre" dans les fichiers, qui augmente fortement lors d'un chiffrement malveillant ;
- **Manipulation de fichiers** : suivi des suppressions, renommages et modifications suspectes ;
- **Types de fichiers ciblés** : focus sur les extensions sensibles (.doc, .pdf, .zip) et les exécutables ;
- **Propagation** : analyse de la dispersion dans les dossiers révélant les "clusters" d'activité typiques des ransomwares ;
- **Informations système** : identifiant du programme, nombre de processus et label de classification (ransomware ou non).

Notre dataset de référence contient 268 748 enregistrements de comportements normaux, servant de base pour détecter les écarts.

Génération de données synthétiques : préparer l'inconnu

Pour dépasser les limites des données réelles — incomplètes, obsolètes et peu variées face aux évolutions rapides des cybermenaces — nous générons des scénarios synthétiques réalistes. Entraînés sur des données authentiques, ces modèles créent des attaques hypothétiques qui :

1. simulent des menaces futures inédites ;
2. combinent des techniques de manière innovante ;
3. testent les frontières de la détection.

Nous garantissons le réalisme en respectant une logique temporelle, des relations naturelles entre indicateurs (par exemple : hausse d'entropie lors de renommages) et des motifs de propagation crédibles. Cette approche transforme la défense en stratégie proactive, renforçant la résilience des systèmes contre les ransomwares de demain.

Autres informations

— Langue : Français
— TLP : VERTE
— Enregistrement / Rediffusion : Autorisé
— Durée : 30 minutes