

Introduction

Le développement sécurisé est une tâche ardue. Si on porte notre regard vers le langage de programmation C, un guide [Can14]¹ porté par l'INRIA² est complet en 133 pages tandis qu'un guide pour du développement sécurisé [ANS20] produit par l'ANSSI³ comprends 182 pages. Cette comparaison met en évidence la discipline requise par le développeur pour faire de la programmation sécurisée ; en sus des connaissances, pour améliorer son efficacité, en cryptologie, en architecture matérielle et en programmation bas niveau .

Malheureusement, malgré ces compétences, des erreurs peuvent être produites puis exploiter pour réaliser des attaques sur ces systèmes sécurisés. Il existe de nombreuses classes d'attaques, certaines exploitant les défauts de conception (type A) tandis que d'autres utilisent les caractéristiques matériels (type B). Pour limiter ces effets de bords, la pratique de la programmation formelle permet de contraindre le développeur et empêcher l'apparitions de ces erreurs. La production de preuve formelle du code à l'issue de cet exercice permet d'avoir des garanties contre les attaques de type A.

En revanche, pour se défendre d'attaques de type B (ou attaques par canal auxiliaire) dépendantes du matériel support du programme, il est plus difficile d'avoir une méthode miracle. Actuellement, la solution la plus courante est d'identifier les attaques existantes pour ajouter les contre-mesures adéquates permettant d'avoir un système sécurisé. Une sous-classe d'attaque continue malgré tout de résister à cette méthode : les attaques temporelles.

Découverte par Paul Kocher en 1996 [Koc96], il les décrit comme «une mesure précise du temps requis par des opérations sur les clés secrètes, permettrait à un attaquant de casser le cryptosystème». Face à cette menace, l'enjeu d'avoir un code *achrognostique*⁴ vient se rajouter aux pratiques de programmations sécurisées. Et pourtant, si contre les attaques de type A on arrive à concevoir des preuves mathématiques de sécurité associées à nos systèmes sécurisés, les garanties contre les attaques de type B sont plus faible ou inexistente.

En 2024, les travaux de SCHNEIDER et al. [Sch+24] prouvent qu'un usage inadéquat de compilateur sur un système sécurisé introduit des failles exploitables. Ces résultats, observables partiellement avec des travaux antérieurs (par exemple [DBR19]), montrent qu'un usage inadéquat d'options fournies au compilateur optimise un code prouvé sécurisé et retire les protections indiquées par le développeur. Cela nous amène à plusieurs questions de recherche (QR) que nous tenterons de répondre à travers ce document.

QR1 Est-il possible de détecter les failles qui permettent une attaque temporelles ?

QR2 Est-il possible d'automatiser la détection de ces failles ?

QR3 Est-il possible d'étendre ce mécanisme entre différentes architectures ?

Les réponses à ces questions permettraient de développer des systèmes sécurisés, communs entre différents supports et d'avoir des garanties de sécurité.

1. Développé par Anne Canteaut, chercheuse de l'équipe COSMIQ, récemment entrée à l'Académie des Sciences

2. Institut National de Recherche en Informatique et Automatique

3. Agence nationale de la Sécurité des Systèmes d'Information

4. Néologisme de Thomas Pornin dans son article *Constant-Time Code : The Pessimist Case* [Por25] pour désigner un code sans connaissance de temps

Fin d'introduction - à finir

Dans la première section nous reviendrons sur les attaques temporelles, leurs impacts et comment s'en protéger. Puis, Dans la deuxième section nous présenterons les outils disponibles à l'analyse et pour la détection de failles. Nous continuerons, dans la troisième section, avec la présentation de nos contributions. *Enfin, dans la quatrième section nous présenterons les mécanismes présent au plus bas niveau de l'informatique pour se protéger des attaques temporelles.*

Ce travail a été réalisé au sein du centre INRIA de Paris dans le cadre du projet Everest concernant la mise au point de HACL*.

Préambule

HACL*⁵

Acronyme pour "High assurance cryptography library", lire "*HACL star*". Il s'agit d'une bibliothèque cryptographique développée au sein du **Projet Everest**⁶. Initié en 2016, ce projet porté par des chercheurs de l'INRIA (équipe PROSECCO⁷), du Centre de Recherche Microsoft et de l'Université Carnégie Mellon a pour but de concevoir des systèmes informatiques formellement sécurisés appliqués à l'environnement HTTPS. Cette bibliothèque écrite en F* ("F star") implémente tous les algorithmes de cryptographie modernes et est prouvée mathématiquement sûre. Elle est ensuite transcrite en C pour être directement employée dans n'importe quel projet. HACL* est notamment utilisé dans plusieurs systèmes de production, notamment Mozilla Firefox, le noyau Linux, le VPN WireGuard, et bien d'autres *etc.*

Binsec⁸

Binary Security est un ensemble d'outils open source développé pour améliorer la sécurité des logiciels au niveau binaire. Ce logiciel est développé et maintenu par une équipe du CEA List de l'Université Paris-Saclay, et accompagné par des chercheurs de Verimag⁹ et de LORIA¹⁰. Il est utilisé pour la recherche de vulnérabilités, la désobfuscation de logiciels malveillants et la vérification formelle de code assembleur. Grâce à l'exécution symbolique et l'interprétation abstraite, Binsec peut explorer et modéliser le comportement d'un programme pour détecter des erreurs; détection réalisée avec des outils de fuzzing et des solveurs SMT.

5. <https://hacl-star.github.io/>

6. <https://project-everest.github.io/>

7. Équipe de recherche rattachée au centre INRIA de Paris, focalisée sur les méthodes formelles et la recherche en protocoles cryptologiques. Pour ces objectifs, l'équipe développe des langages de programmation, des outils de vérification...

8. <https://binsec.github.io/>

9. Verimag est un laboratoire spécialisé dans les méthodes formelles pour une informatique sûre, avec des applications aux systèmes cyber-physiques. Fondé en 1993 au sein de l'Université Grenoble Alpes, puis rejoint par le CNRS, il a pour objectif la sécurité dans les domaines des transports et de la santé.

10. Laboratoire lorrain de recherche en informatique et ses applications; créé en 1997, c'est un centre de recherche commun au CNRS, l'Université de Lorraine, CentraleSupélec et l'Inria.

Présentation, enjeux et attaques

Ce premier chapitre a pour but de présenter les enjeux de la sécurité informatique face aux attaques par canal auxiliaire et d'introduire les attaques temporelles. Nous distinguerons les attaques par canal auxiliaire en deux catégories, montrant ainsi la diversité et les potentiels dangers pour un système sécurisé ignorant de cette menace.

1.1 L'exécution du code est observable...

L'Informatique repose sur deux fondations que l'on tend à distinguer dans l'enseignement : le matériel et le logiciel. Pourtant, si on gardait séparé ces deux domaines, on aurait des tas de piles de métal et de plastiques ou des bibliothèques de livres plein d'idées intéressantes. Au contraire, combiner les deux parties permet de réaliser des prouesses technologiques et scientifiques. Ainsi, lorsque l'on conçoit un système sécurisé, il faut prendre en compte ces deux composantes. Or pour implémenter un système sécurisé, il ne faut pas seulement un logiciel sécurisé, il est aussi important que le matériel le soit. Oublier comment fonctionne un support informatique, c'est oublier que programmer se résume à manipuler de l'électricité.

Les attaques par canal auxiliaires consistent à exploiter les caractéristiques matériels du support pour gagner en connaissances sur le programme ciblé. Puis exploiter ces connaissances pour acquérir d'avantages d'informations privées : identifiants, clés secrètes, messages personnels. On leur attribue le terme "canal auxiliaire" car il ne s'agit pas d'essayer de pousser dans ses limites un logiciel ou vérifier que tous les cas particulier sont gérés à travers le canal conçus par le développeur (une interface graphique souvent) mais plutôt de se positionner hors du cadre. Voici quelques travaux présentant une attaque par canal auxiliaire et surtout le canal exploité :

- [KJJ99] Consommation d'énergie
- [AKS06] Prédiction de branchement
- [Mas+15] Variation de température
- [Pes+16] Accès à la mémoire DRAM

Le point commun de ces attaques est la nécessité d'avoir un point de contact avec la cible. Il faut que l'attaquant puisse récupérer le matériel informatique ou le programme qu'il souhaite attaquer pour ensuite poser des sondes/capteurs enfin d'accumuler de la connaissance et monter son exploitation.

Une autre technique d'attaque consiste à venir introduire une erreur dans le déroulement normal d'un programme. Il s'agit d'une attaque par injection de faute. Originellement [Avi71] les fautes étaient "naturelles" : un défaut dans le code, un problème avec la transcription vers du code machine, un défaut d'un composant dans le système ou une interférence. Ces interférences sont causées par une irrégularité de l'alimentation électrique, des radiations électromagnétique, une perturbation environnementale *etc* ... En 2004, BAR-EL et al. dans leur article *The Sorcerer's Apprentice Guide to Fault Attacks* [Bar+04] effectuent

un tour d’horizon des techniques, montrant l’efficacité de cette méthode sur RSA¹, NVM², DES³, EEPROM⁴, JVM⁵. On y retrouve enfin une liste de contre-mesures et de méthodes de protections contre ces attaques.

Ainsi, donner un accès physique à un inconnu est une porte d’entrée pour un attaquant. Pourtant, penser que l’accès physique au support est une condition nécessaire et suffisante pour réaliser une attaque par canal auxiliaire est une erreur.

1.2 ...à distance

En effet, il est possible de réaliser des attaques à distance en exploitant d’autres failles de sécurité d’un programme ou d’autres caractéristiques matériels. L’attaque présentée par LIU et al. dans “ Last-Level Cache Side-Channel Attacks are Practical ” [Liu+15] repose sur la conception des services clouds où les machines virtuelles accèdent au même matériel. Tandis que la virtualisation crée l’illusion de compartimentation entre les sessions, en réalité, les adresses mémoires pointent vers une ressource physique partagée. Ainsi, l’exploitation du cache du dernier niveau (LLC) permet à un co-hôte de récupérer les clés secrètes d’un autre utilisateur. L’attaquant remplit le cache, puis mesure les temps d’accès vers ces registres. Si des modifications apparaissent dans ces temps, cela signifie que la victime a accédé à ces registres. En répétant cette opération, l’attaquant peut reconstruire les clés secrètes de la victime.

D’autres attaques distantes comme celle de LIU et al. existent [YGH16 ; Mog+17 ; VPS18], mais on observe rapidement que ces techniques emploient aussi la méthode de chronométrage. En effet, si on cible un algorithme et que l’on mesure son temps d’exécution. Si en fournissant différentes entrées (que l’on considère secrètes) des variations sont observées entre les mesures, alors cela signifie que l’algorithme présente une dépendance à ces entrées. Généralement une sous-fonction de cet algorithme est responsable de ces variations. Cette classe d’attaque est appelée «*attaque temporelle*»⁶.

Le lien entre temps et exécution de code est connu depuis le début de l’informatique. Le temps est le marqueur de performance, d’efficacité d’un programme. En revanche, l’idée d’exploiter cet indice pour réaliser une attaque est arrivée plus tardivement. KOCHER nous présente le premier, en 1996, comment monter une attaque en utilisant ce canal.

Ce lien entre temps et exécution est connu, pourtant la mesure de l’ampleur de la fuite d’information transmise par ce canal n’est pas triviale ; ni à son époque, ni à celle-ci.

```

1  bool check_pwd(msg, pwd){
2      if (msg.length != pwd.length){
3          return False
4      }
5      for(int i = 0; i < msg.length; i++){
6          if(msg[i] != pwd[i]){
7              return False
8          }
9      }
10     return True
11 }
```

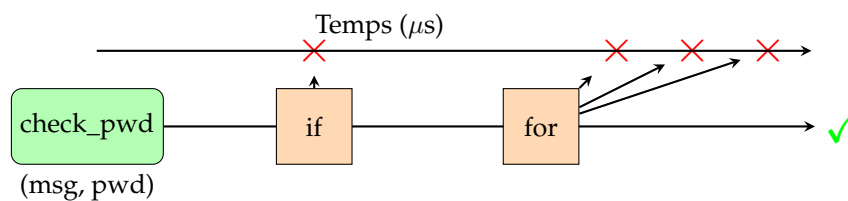
Code 1 – Exemple de code vulnérable à une attaque temporelle

-
1. Chiffrement asymétrique par clés secrète du nom de ces auteurs. Standardisé en 1983.
 2. Non Volatile Memory ou mémoire non volatile est un composant informatique qui conservent son contenu en l’absence d’électricité.
 3. Algorithme de chiffrement symétrique par bloc. Standardisé en 1977
 4. Electrically-Erasable Programmable Read-Only Memory ou mémoire morte effaçable électriquement et programmable.
 5. Machine virtuelle qui exécute des programmes compilés en bytecode Java.
 6. Le terme générique dans la recherche scientifique est «*time attack*». Une traduction plus précise serait «*attaque par chronométrage*». on choisit ici d’utiliser le terme «*attaque temporelle*» car il est moins lourd et renvoie directement vers la faille exploitée plutôt que sur la méthode employée.

Si nous prenons le code présenté par le code 1, on peut observer que la fonction `check_pwd` compare deux chaînes de caractères. Si elles sont de même longueur, elle les compare caractère par caractère. Si elles sont de longueurs différentes, la fonction retourne immédiatement `False`. Ainsi, si l'on fournit un mot de passe de longueur différente, le temps d'exécution sera constant et court. En revanche, si l'on fournit un mot de passe de même longueur, le temps d'exécution dépendra du nombre de caractères identiques entre les deux chaînes. En effet, la fonction s'arrêtera dès qu'un caractère différent est trouvé. Ainsi, en mesurant le temps d'exécution pour différents mots de passe, un attaquant peut déduire des informations sur le mot de passe correct.

On peut synthétiser les exécutions de la fonction `check_pwd` en un graphe comme celui présenté par la figure 1.1. Chaque interruption de la fonction peut être observée et mesurée, permettant ainsi de régénérer le mot de passe. Bien sûr la connaissance du protocole cible est requise ou alors il faut réaliser un travail de rétro-ingénierie pour calibrer l'attaque.

FIGURE 1.1 – Suivi du temps d'exécution pour différents mots de passe



Cette méthode est plus efficace qu'une attaque par force brute. En effet, si l'on suppose que le mot de passe est de 8 caractères de l'alphabet latin. On a alors 256 possibilités par caractère, pour un total de $256^8 = 2^{64}$ possibilités. En revanche, si l'on utilise la méthode de l'attaque temporelle, on peut réduire le nombre de possibilités à $8 + 8 \times 256 = 2056$ possibilités. En effet, on cherche dans un premier temps à identifier la longueur du mot de passe, puis on identifie caractère après caractère pour trouver le bon secret. Avec des temps d'exécution court on est dans les cas de figure d'échec, tandis qu'avec un allongement du temps d'exécution on sait que l'on est sur la bonne piste.

Les attaques temporelles présentent la particularité d'être générique. Tandis que les attaques décrites précédemment nécessite des conditions d'accès ou d'initialisation plus importante, cette classe d'attaque présente l'avantage d'être réalisable sur tous les types de systèmes, et notamment les systèmes accessible par internet. La connaissance de cette menace est donc primordiale pour l'implémentation et la mise en service de produit sur internet.

Par la suite du document, le terme "fuite" sera utilisé pour désigner un extrait du programme qui peut être exploité pour réaliser une attaque temporelle. Si on reprend le code 1, les branchement conditionnels ligne [4,6] sont des fuites d'informations. C'est grâce à ces instructions que l'attaque décrite précédemment est réalisable.

Nous allons maintenant nous intéresser aux moyens et méthodes à notre disposition pour se protéger contre les attaques temporelles.

Protection

Ce deuxième chapitre montre les innovations nécessaires pour se protéger des attaques temporelles. On y découvre les bonnes pratiques de programmation, les premiers outils automatique de vérification de code ainsi que les limitations auxquelles est confronté le développeur qui souhaite être résistant à ces attaques.

2.1 Bonne pratique et usages

Face à la menace des attaques temporelles, quelles solutions peuvent être mises en place pour protéger nos systèmes informatiques? Cette attaque a besoin d'un accès au système et d'un chronomètre. Comme on est dans un contexte de systèmes accessibles par internet, altérer ou retirer l'accès signifie perdre en qualité ou supprimer le service proposé. Il faut donc que notre approche cible plutôt l'utilisation du chronomètre.

Il faut donc programmer de tel sorte que sur toutes les entrées possibles de notre système informatique aucune variation de temps ne peut être observée entre les exécutions. Trois méthodes existent pour pallier à ce problème.

Programmation en temps constant

La programmation en temps constant ou «*Constant-Time Programming*», est une pratique de programmation qui vise à résoudre exactement ce problème. Directement lié à la complexité algorithmique, cette pratique modifie et adapte les algorithmes pour que toutes les opérations effectuées aient un temps d'exécution identique.

PORNIN [Por16] présente tous les éléments à adapter pour configurer un code respectant la politique de programmation en temps constant. Si les opérations élémentaires respectent "naturellement" cette politique; les **accès mémoires**, les **sauts conditionnels**, les **sopérations de décalages/rotations** et les **divisions/multiplications** sont les opérations à adapter en fonction de la plateforme cible. Les descriptions rapportées ci-dessous sont issues de [Por16].

Accès mémoire

Un chargement depuis la mémoire d'une information est une source de variation. On a vu précédemment [Liu+15; Pes+16] que l'usage d'un cache mémoire est un canal d'accès pour réaliser une attaque. En effet, l'utilisation d'un cache permet de distinguer les appels entre les données déjà mises en mémoire ou pas. De plus, les changements de valeur dans celui-ci peuvent aussi être observé après exécution.

Décalage et rotation

Ces opérations binaires sont ou ne sont pas en temps constant en fonction des CPU sur lequel le code est exécuté. Certains ont un "barrel shifter" qui permet d'effectuer directement les instructions correspondantes. Cela impacte directement les algorithmes dépendant de décalages logiques comme le chiffrement RC5.

Saut conditionnel

Les sauts conditionnels sont des instructions qui, comme pour les accès mémoires, demandent de charger les adresses des instructions suivantes. Or, comme un compilateur tend à précharger les instructions suivantes, il va charger les deux côtés du saut conditionnel puis defausser la branche inutile ; ce qui entraîne un léger ralentissement. En revanche, il est important de noter que si le branchement est indépendant d'une variable secrète, il n'est pas nécessaire de le modifier. Par exemple si j'ai un compteur et que mon programme doit terminer après un certain nombre d'itérations, aucune fuite ne sera observée.

Division

Certaines architectures ont des instructions de divisions spécifiques qui permettent d'accélérer le calcul, les autres emploient des sous-programmes dédiées souvent optimisés en opération de masquage et de décalage. La norme C entraîne elle aussi de la confusion car elle impose $(-1)/2 = 0$; il faut donc être familier avec les spécificités du processeur pour affiner l'usage de cette opération.

Multiplication

Enfin, la multiplication, elle aussi dépendante des variables d'entrées, présente une fuite d'information importante. Mais les CPU les plus récents (rédigé en 2016) ont implémenté cette opération en temps constant. Cela suit l'évolution des compilateurs et des processeurs qui tendent à accélérer les opérations et réduire le nombre d'instruction total.

En reprennant ces règles, on peut modifier notre exemple de code 1 et appliquer des modifications sur lignes que l'on a déjà ciblées comme fuites d'informations. Les modifications sont libre au choix du concepteur. Voici une correction qui peut être réalisée :

```

1  bool check_pwd(msg, pwd) {
2      // Hachage
3      char msg_hash[SHA256_DIGEST_LENGTH]; sha256_hash_string(msg, msg_hash);
4      char pwd_hash[SHA256_DIGEST_LENGTH]; sha256_hash_string(pwd, pwd_hash);
5
6      // Comparaison
7      bool equal = true;
8      for (int i = 0; i < SHA256_DIGEST_LENGTH; i++) {
9          if (msg_hash[i] != pwd_hash[i]) {
10             equal = equal && false;
11          } else {
12             equal = equal || false;
13          }
14      }
15      return equal;
16  }

```

Code 2 – Exemple de correction pour rendre un code résistant aux attaques temporelles

On voit que le premier branchement a été remplacé par un hachage des paramètres d'entrées. Cette opération est considérée ici en temps constant mais peut ne pas l'être. Il faut être vigilant sur toutes les briques d'algorithme que l'on souhaite utiliser. Enfin, le second branchement conditionnel est purement supprimé, le parcours des tableaux se fait entièrement.

Avec cette modification, on a un code 2 qui ne présente plus de fuite de données. Pourtant, on peut avoir un doute sur l'usage de la fonction "*sha256_hash_string*". Si cette fonction n'est pas elle même implémentée selon la politique temps constant, on a alors introduit une nouvelle surface de fuite d'informations. Il faut vérifier notre code pour supprimer ce doute.

Outils de garanties

Plusieurs outils existent et peuvent être utilisés tous au long du processus de développement d'un système sécurisé. Cela peut être durant la phase de conception du code source, au moment de la compilation ou encore en vérification de la compilation.

Une solution légère est de se servir du système libre «**Compiler Explorer**¹». Avec à disposition un éditeur de texte, il est possible de voir comment sera généré le code assembleur. En reprenant une partie du code 1.1, on peut voir sur la figure 2.1 que le choix du compilateur, ici sa version, introduit une légère modification. Ce changement n'est pas perceptible sans observation directe ce perçoit directement grâce à la petite taille du code observé.

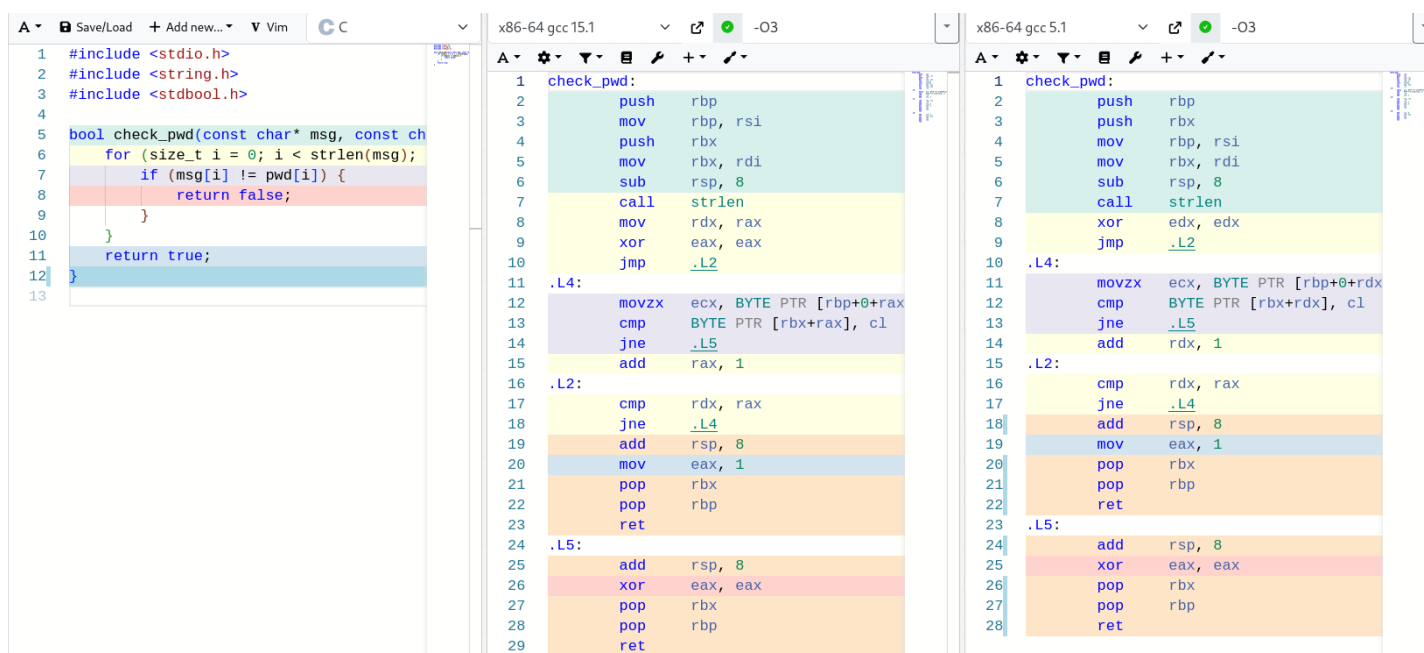


FIGURE 2.1 – Capture d'écran de comparaison de code assembleur x86_64 entre GCC 15.1 et GCC 5.1

Si l'on souhaite faire une analyse à l'échelle d'un projet, ce parcours à la main des fonctions ou de morceaux de fonctions est réellement fastidieux. Il faut mieux déléguer ce travail à un outil conçu pour vérifier la présence de fuite.

Plusieurs articles référencent l'ensemble des outils existant [Jan+21; Gei+23] pour réaliser ce travail. Le tableau 2.1 de JANCAR et al. liste 24 outils en libre accès conçus pour détecter des failles par canal auxiliaire.

Ils sont listés alphabétiquement et sont précisés le type de fichier analysé (*Cible*), la méthode d'analyse réalisée (*Techn.*) et les garanties attendues de ces analyses (*Garanties*). On reviendra plus en avant avec ces détails de méthodes et de fonctionnement dans le chapitre ??.

1. <https://godbolt.org/>

TABLE 2.1 – Liste d’outils de vérification, source [Jan+21]

Cible : [C, Java] = Code source, Binaire = Binaire, DSL = Surcouche de langage, Trace = Trace d’exécution, WASM = Assembleur web.

Techn. : Formel = Programmation formelle, [Symbolique, Dynamique, Statistique] = type d’analyse.

Garanties (*Sécurité face aux attaques temporelles*) : ● = Analyse correct, ▲ = Correct mais avec des limitations, ○ = Aucune garantie,

★ = Vérification d’autres propriétés.

Outil	Cible	Techn.	Garanties
ABPV13 [Alm+13]	C	Formel	●
Binsec/Rel [DBR19]	Binaire	Symbolique	▲
Blazer [Ant+17]	Java	Formel	●
BPT17 [BPT17]	C	Symbolique	▲
CacheAudit [Doy+13]	Binaire	Formel	★
CacheD [Wan+17]	Trace	Symbolique	○
COCO-CHANNEL [Bre+18]	Java	Symbolique	●
ctgrind [Lan10]	Binaire	Dynamique	▲
ct-fuzz [HEC20]	LLVM	Dynamique	○
ct-verif [Bar+16]	LLVM	Formel	●
CT-WASM [Wat+19]	WASM	Formel	●
DATA [Wei+20; Wei+18]	Binaire	Dynamique	▲
dudect [RBV17]	Binaire	Statistique	○
FaCT [Cau+19]	DSL	Formel	●
FlowTracker [RPA16]	LLVM	Formel	●
haybale-pitchfork [Dis20]	LLVM	Symbolique	▲
KMO12 [KMO12]	Binaire	Formel	★
MemSan [Tea17]	LLVM	Dynamique	▲
MicroWalk [Wic+18]	Binaire	Dynamique	▲
SC-Eliminator [Wu+18]	LLVM	Formel	●
SideTrail [Ath+18]	LLVM	Formel	★
Themis [CFD17]	Java	Formel	●
timecop [Nei18]	Binaire	Dynamique	▲
VirtualCert [Bar+14]	x86	Formel	●

Une dernière solution serait d’utiliser un compilateur spécialisé qui produit un code assembleur sans fuite [Bor+21; RLT15] ou d’utiliser un compilateur formel comme *CompCert* [Ler+05]. Cette solution rencontre en pratique de nombreux problèmes que l’on se garde pour la section 2.2 Limitations.

Écriture en code assembleur

Enfin, la dernière méthode pour obtenir un code sécurisé et sans fuite c’est de programmer directement en assembleur. De cette manière on a un contrôle total sur le flot d’exécution de notre programme, on peut ainsi insérer des optimisations qu’un compilateur pourrait ignorer. Écrire en assembleur requiert de connaître la plupart des opérandes disponible pour l’architecture que l’on cible et les modèles des composants présent sur le support. Cela nous amène directement aux limitations induites par cette solution.

2.2 Limitations

Écrire en assembleur c’est écrire spécifiquement pour une architecture de processeur. Il faut connaître les instructions adéquates, les potentielles optimisations qui existent sans parler de la syntaxe particulière qui rend son développement plus lent. Travailler en assembleur c’est limiter la portabilité du code proposé or l’objectif derrière le développement d’une librairie sécurisée est de pouvoir être employée par le plus de configuration possibles pour se protéger d’attaques.

Face à cette situation, on choisit donc d’utiliser un compilateur spécialisé ([Bor+21; RLT15]). Et à nouveau on se retrouve limité parce que ces compilateurs ne supportent pas l’ensemble du jeu d’instruction d’une architecture, ont besoin d’instructions supplémentaires (des annotations de code) pour réaliser la compilation, n’implémente pas les optimi-

sations qui apparaissent sur les processeurs les plus récents ou encore ne sont adapté qu'un seul langage de programmation.

À nouveau, on se retrouve donc à utiliser les compilateurs communs GCC et LLVM pour notre solution sécurisé. On se doit donc de programmer en respectant la politique temps constant. Et si cette pratique semble faire ses preuves, on peut lire dans la présentation de l'outil d'analyse Binsec "Binsec/Rel : Efficient Relational Symbolic Execution for Constant-Time at Binary-Level" :

Conclusion - [DBR19]

Nous avons découvert que `gcc -O0` et des optimisations de `clang` introduisent des infractions à la politique temps constant indétectées par les outils antérieurs

Cette annonce a ensuite été prise en compte par SCHNEIDER et al. qui a mené une enquête sur les bibliothèques cryptographiques sécurisées et résistantes aux attaques temporelles : [Sch+24]. La conclusion principale est que les compilateurs modernes sont devenus assez performant pour voir à travers les astuces employées et qu'une mauvaise utilisation d'optimisation implique l'introduction de faille de sécurité.

Voici un exemple communiqué par SCHNEIDER et al. auprès des chercheurs de Hacl*. On peut voir deux fonctions dans le code 3, «*cmovznz4*» et «*FStar_UInt64_eq_mask*». La première appelle la seconde pour générer un masque qui sera ensuite appliqué au entrée de «*cmovznz4*». On a ici une fonction qui agit comme un branchement conditionnel. Si `cin` vaut 1, alors $r = x$ sinon $r = y$.

```

1  #include <stdint.h>
2
3  static inline uint64_t FStar_UInt64_eq_mask(uint64_t a, uint64_t b)
4  {
5      uint64_t x = a ^ b;
6      uint64_t minus_x = ~x + (uint64_t)1U;
7      uint64_t x_or_minus_x = x | minus_x;
8      uint64_t xnx = x_or_minus_x >> (uint32_t)63U;
9      return xnx - (uint64_t)1U;
10 }
11
12 void cmovznz4(uint64_t cin, uint64_t *x, uint64_t *y, uint64_t *r)
13 {
14     uint64_t mask = ~FStar_UInt64_eq_mask(cin, (uint64_t)0U);
15     uint64_t r0 = (y[0U] & mask) | (x[0U] & ~mask);
16     uint64_t r1 = (y[1U] & mask) | (x[1U] & ~mask);
17     uint64_t r2 = (y[2U] & mask) | (x[2U] & ~mask);
18     uint64_t r3 = (y[3U] & mask) | (x[3U] & ~mask);
19     r[0U] = r0;
20     r[1U] = r1;
21     r[2U] = r2;
22     r[3U] = r3;
23 }

```

Code 3 – Fonction de masquage issu de Hacl*

Avec le compilateur RISC-V `rv64gc clang 15.0.0`, si on entre les options de compilation `-O0` ou `-O1` on peut observer différents résultats. Le plus notable ici est l'apparition de l'instruction `beqz`, qui est un branchement conditionnel, ainsi que la suppression de la fonction de masquage «*FStar_UInt64_eq_mask*». Les optimisations appelées par l'option `-O1` permettent d'identifier le tour de passe passe qui lui été proposé. Le code 2.2a suivent les instructions précisées par le code source, le compilateur avec cette optimisation compile vite. Au contraire du code 2.2b produit par une analyse plus longue du compilateur où les sauts succesifs entre les `beqz` permettent une exécution plus rapide. Les options de compilations sont rapportées en annexe 5.1².

2. <https://gcc.gnu.org/>

<pre> 1 cmovznz4: 2 ... 3 li a1, 0 4 call FStar_UInt64_eq_mask 5 not a0, a0 6 sd a0, -56(s0) 7 ld a0, -40(s0) 8 ld a0, 0(a0) 9 ld a2, -56(s0) 10 and a0, a0, a2 11 ld a1, -32(s0) 12 ld a1, 0(a1) 13 not a2, a2 14 and a1, a1, a2 15 or a0, a0, a1 16 sd a0, -64(s0) 17 ... 18 ret 19 20 FStar_UInt64_eq_mask: 21 addi sp, sp, -64 22 sd ra, 56(sp) 23 sd s0, 48(sp) 24 addi s0, sp, 64 25 sd a0, -24(s0) 26 sd a1, -32(s0) 27 ld a0, -24(s0) 28 ld a1, -32(s0) 29 xor a0, a0, a1 30 sd a0, -40(s0) 31 ld a1, -40(s0) 32 li a0, 0 33 sub a0, a0, a1 34 sd a0, -48(s0) 35 ld a0, -40(s0) 36 ld a1, -48(s0) 37 or a0, a0, a1 38 sd a0, -56(s0) 39 ld a0, -56(s0) 40 srli a0, a0, 63 41 sd a0, -64(s0) 42 ld a0, -64(s0) 43 addi a0, a0, -1 44 ld ra, 56(sp) 45 ld s0, 48(sp) 46 addi sp, sp, 64 47 ret </pre>	<pre> 1 cmovznz4: 2 mv a5, a1 3 beqz a0, .LBB0_2 4 mv a5, a2 5 .LBB0_2: 6 beqz a0, .LBB0_5 7 addi a6, a2, 8 8 bnez a0, .LBB0_6 9 .LBB0_4: 10 addi a4, a1, 16 11 j .LBB0_7 12 .LBB0_5: 13 addi a6, a1, 8 14 beqz a0, .LBB0_4 15 .LBB0_6: 16 addi a4, a2, 16 17 .LBB0_7: 18 ld a7, 0(a5) 19 ld a5, 0(a6) 20 ld a6, 0(a4) 21 beqz a0, .LBB0_9 22 addi a0, a2, 24 23 j .LBB0_10 24 .LBB0_9: 25 addi a0, a1, 24 26 .LBB0_10: 27 ld a0, 0(a0) 28 sd a7, 0(a3) 29 sd a5, 8(a3) 30 sd a6, 16(a3) 31 sd a0, 24(a3) 32 ret </pre>
---	--

(a) Option -O0

(b) Option -O1

FIGURE 2.2 – Comparaison du code 3 en fonction de différentes options de compilation données au compilateur, réalisée avec l'aide de *Compiler Explorer*.

Transition

Implémentations pour un usage industriel

intro

3.1 Identification des besoins et spécificités

On a pu voir grâce aux chapitres précédents que la conception et l'implémentation d'un système sécurisé est un problème difficile. Une première étape est de concevoir des primitives et protocoles mathématiquement sécurisés. Une seconde étape est de s'assurer que leurs implémentations sont effectivement sécurisées, d'abord d'un point de vue mathématique contre des attaques logiques (aspect fonctionnel : le code implémente correctement les bons concepts cryptographiques), mais aussi contre des attaques très bas niveau, les attaques temporelles.

Avec l'objectif de concevoir un système sûr, il nous faut donc identifier toutes les tâches à réaliser pour arriver à bout de ce projet. En plus de ce travail de planification, l'identification et l'intégration d'outils déjà implémentés nous permettra de d'avancer plus rapidement vers cet objectif.

Point de départ

En reprenant ces deux étapes, on va identifier quels sont nos leviers et nos possibilités pour un développeur pour avancer dans la conception de notre graal.

La première étape de conception de primitives cryptologiques et de protocole n'est pas du ressort du développeur. Elle appartient aux cryptologues et aux chercheurs en sécurité mathématique. Ce sont eux qui conçoivent et maintiennent des bibliothèques cryptographiques, des boîtes à outils qui proposent les briques de sécurité nécessaires aux systèmes sécurisés.

Plusieurs bibliothèques existent [AHa98 ; Por16 ; Pol+20] et remplissent différents objectifs : rétro-compatibilité, politique temps constant, *etc.* Notre choix est à réaliser en fonction des spécificités du produit que l'on cherche à déployer.

La seconde étape est à distinguer en deux parties. Cette opération de vérification de la sécurité de l'implémentation peut-être réalisée sur le produit fini et sur les bibliothèques employées par le produit. Comme introduit, cette étape a pour objectif la vérification formelle du code du programme et la vérification matérielle au niveau assembleur.

Utiliser la bibliothèque **Hacl*** [Pol+20 ; Zin+17] permet d'avancer la première étape et la première partie de la seconde étape. Cette bibliothèque a été conçue formellement et vient avec les preuves mathématiques de la sécurité de son implémentation. Comme présenté en Préambule, cette bibliothèque est programmée en F*. Le projet permet une exploitation en C et en assembleur [Zin+17].

En revanche, la seconde étape de la seconde partie nous demande une vérification au niveau de l'assembleur. Si certaine partie de cette librairie sont codées en assembleur, la majorité du projet reste du F* traduit vers C. Il faut réaliser une analyse. Dans le cadre de cette étude, l'outil d'analyse binaire retenu pour réaliser cette tâche est **Binsec**. Cette outil est implémenté en Ocaml et est maintenu par une équipe de chercheurs ingénieurs géographiquement proche de l'équipe PROSECCO Inria. Cet avantage permet des échanges plus directs et donc une facilité quand à la mise en place du projet.

L'objectif est donc d'analyser Hacl* dans son entièreté. Avec cette analyse complète, si elle est correcte, alors les deux étapes de réalisation d'un système sûr seront réalisées. Cela signifie que la première librairie cryptographique formellement sûre et résistante aux attaques temporelles sera conçus.

Objectifs à réaliser

Sans reprendre les explications du fonctionnement de Binsec, voir "[ref vers fonctionnement de Binsec](#)", l'analyse se réalise sur un fichier binaire à l'aide d'un carnet d'instructions à préciser. Avec ce point de départ, on peut commencer à construire notre carnet de spécifications.

Fichier binaire. Il faut donc des fichiers binaires à fournir à Binsec. Or comme chacun le sait, plus un binaire est imposant, plus son analyse est difficile. Et comme Binsec emploie l'analyse symbolique, explorer un binaire imposant a un coût de mémoire quadratique sur le parcours des instructions du binaire. L'idéal est donc d'analyser plein de petits fichiers binaires.

Analyse complète. Chaque fonction de Hacl* doit être analysée. En poursuivant la condition précédente, on peut essayer de concevoir un binaire par fonction analysé. On distribue ainsi l'analyse et on parcourt ainsi toute les fonctions présentent dans la librairie.

Analyse correcte. Si on se rappelle comment fonctionne les optimisations (voir le tableau 5.1) il faut faire attention avec certaines optimisations qui simplifient le code par soustraction d'opérations. Le fichier ne doit pas seulement contenir un appel de fonction, il faut une légère mise contexte.

```

1  #include <stdlib.h>
2
3  #include "Hacl_AEAD_Chacha20Poly1305_Simd128.h"
4
5  #define BUF_SIZE 16384
6  #define KEY_SIZE 32
7  #define NONCE_SIZE 12
8  #define AAD_SIZE 12
9  #define TAG_SIZE 16
10
11 uint8_t plain[BUF_SIZE];
12 uint8_t cipher[BUF_SIZE];
13 uint8_t aead_key[KEY_SIZE];
14 uint8_t aead_nonce[NONCE_SIZE];
15 uint8_t aead_aad[AAD_SIZE];
16 uint8_t tag[TAG_SIZE];
17
18 int main (int argc, char *argv[])
19 {
20   Hacl_AEAD_Chacha20Poly1305_Simd128_encrypt
21     (cipher, tag, plain, BUF_SIZE, aead_aad, AAD_SIZE, aead_key, aead_nonce);
22   exit(0);
23 }
```

Code 4 – Code d'anlayse de la fonction Hacl_AEAD_Chacha20Poly1305_Simd128_encrypt, testé lors de la prise main de Binsec et Hacl*

De même, comme nos fichiers analysés font appel à la librairie extérieur Hacl*, l'emploi de l'option `-static` est nécessaire pour prévenir la mise place de lien vers la librairie par-

tagée dans le fichier binaire. Cette option ne nuit pas à la qualité de l'analyse, elle permet en revanche d'avoir tous les éléments sous la main lorsque l'on déassemble un fichier binaire. Retirer cette option lors de la compilation, c'est se rajouter des lourdeurs et rallonger la temps requis pour la vérification manuelle d'un fichier.

Couverture de compilateur. Les travaux de SCHNEIDER et al. [Sch+24] ont clairement mis en évidence que le choix du compilateur est à considérer. Il faut donc identifier quel compilateur nous permet d'avoir des fichiers binaires les plus sécurisés. On peut aussi identifier quels optimisations produisent la rupture de sécurité dans le binaire en étudiant plus en avant le comportement de ceux-ci.

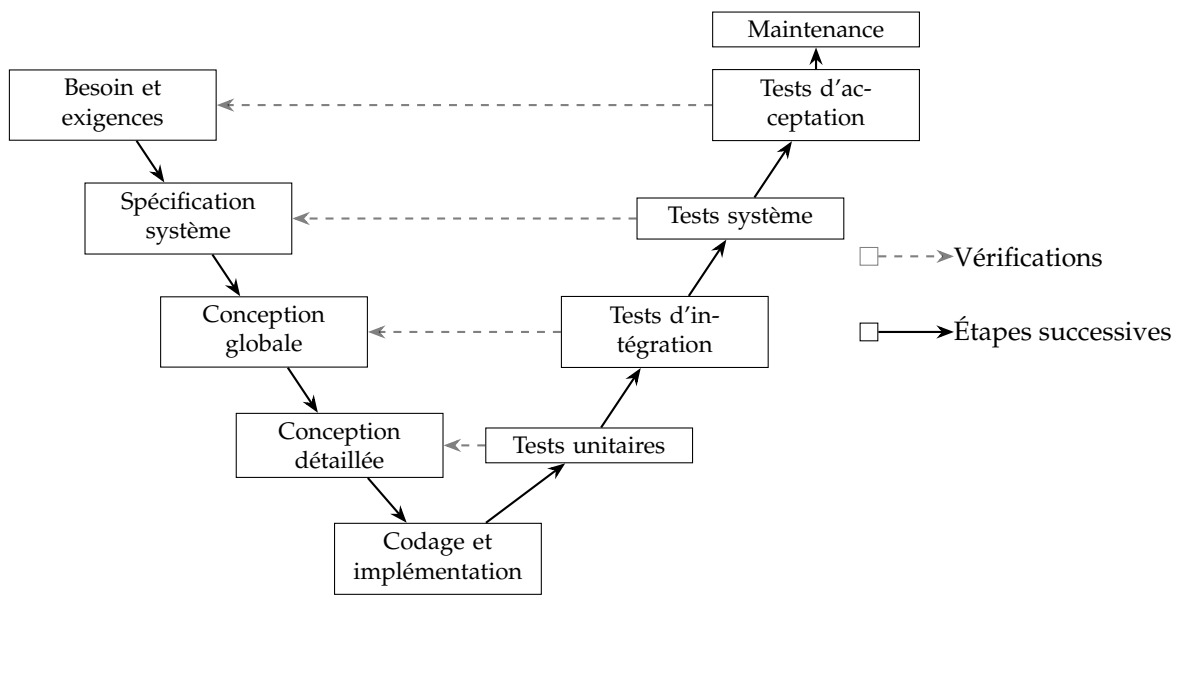
Couverture d'architectures. x86_64 et ARM sont les architectures matérielles les plus répandues dans le monde. Étendre l'analyse vers différentes plateformes et observer les différences qui émergent nous permettraient d'avancer dans la direction de la conception d'une librairie cryptographique universelle. On aussi étendre l'analyse vers d'autres architectures comme PowerPC ou RiscV.

Automatisation. Faire cette analyse sur un fichier binaire, comme le code 4, avec trois axes de complexité (complétude, de la couverture d'architectures et des compilateurs) n'est pas envisageable à la main. Il faut absolument que cette analyse soit automatisée.

3.2 Initialisation et tests variés

Dans le cadre de la programmation sécuritaire, où sont développés les systèmes avec pour objectif de un accident par siècle (métros automatiques, trains, avions...), les projets sont conçus selon le principe du cycle en V. Au contraire de la méthode Agile où on avance vers les problèmes en les résolvants au fur et à mesure, avec ce principe le développement est beaucoup plus long mais permet d'esquiver les problèmes qui, dans son contexte d'usage, peuvent entraîner des décès.

FIGURE 3.1 – Cycle en V



Appliqué cette méthode à l'entièreté de ce projet n'est pas envisageable à cause du coût temporel qui est très élevé. On va se concentrer sur la réalisation d'une preuve de concept et se concentrer sur la partie automatisation. La conception du produit sera minimale et le développement des couvertures sera soumis à un futur travail.

Identification des besoins et exigences

On a déjà conçus notre carnet d'exigence, en revanche on ne connais pas le comportement des outils que l'on souhaite employer. La première opération est de s'approprier le

fonctionnement des outils que l'on s'apprête à employer. Le code 4 est un exemple de test réalisé dans cette phase du projet.

Binsec est un outil uniquement utilisable au travers d'un terminal. Il s'invoque avec son alias, le binaire à analyser et les options de l'analyse qui sera effectué :

```
$ binsec -sse -sse-script $(BINSEC_SCRIPT) -checkct $(BINARY)
```

Code 5 – Commande Binsec basique

L'option `-sse` permet d'activer l'analyse par exécution symbolique, `-sse-script` associer à un fichier (ici `BINSEC_SCRIPT`) permet d'instruire notre analyse, préciser des stubs¹ et des initialisations, enfin `-checkct` active la vérification des propriétés temps constant au sein du fichier binaire indiqué par `BINARY`. Binsec renvoie dans le terminal le résultat de son analyse : `[secure, unknown, insecure]`. Le second est invoqué lorsque l'analyse est incomplète.

Cette phase «Test et Identification des exigences» permet de confronter plusieurs fonctions de Hacl* et de se familiariser avec le langage d'instruction qu'admet l'option `-sse-script`. Un tutoriel complet est accessible pour comprendre le fonctionnement l'outil Binsec depuis sa page officielle².

```

1  starting from core with
2      argv<64> := rsi
3      arg1<64> := @[argv + 8, 8]
4      size<64> := nondet                # 0 < strlen(argv[1]) < 128
5      assume 0 < size < 128
6      all_printables<1> := true
7      @[arg1, 128] := 0
8      for i<64> in 0 to size - 1 do
9          @[arg1 + i] := nondet as password
10         all_printables := all_printables && " " <= password <= "~"
11     end
12     assume all_printables
13 end
14
15 replace <puts>, <printf> by
16 return
17 end
18
19 reach <puts> such that @[rdi, 14] = "Good password!"
20 then print ascii stream password
21
22 cut at <puts> if @[rdi, 17] = "Invalid password!"
23
24 halt at <printf>

```

Code 6 – Instructions permettant de trouver le mot d'un passe d'un binaire exercice

Ce code présenté ici est un exemple d'usage de Binsec et permet de réaliser une attaque sur un binaire issu d'une plateforme d'apprentissage à la sécurité logiciel³. L'exercice consistant à retrouver le mot de passe caché d'un binaire. Dans le cadre de notre exercice d'analyse de la politique temps constant, le script 7 est plus simple.

1. Terme anglais du lexique de la rétro-ingénierie ; module logiciel simulant la présence d'un autre.

2. <https://binsec.github.io/>

3. <https://crackmes.one/>

Ce script a été conçu avec pour objectif de vérifier les résultats apportés par [Sch+24] concernant une fuite présente sur la fonction «*FStar_UInt64_eq_mask*» et d'étendre l'analyse vers d'autres architectures. Dans une première démarche d'automatisation, ce code à été généré automatiquement par un script shell. On voit ici que l'analyse ne parcourt pas l'entièreté du binaire, seulement 8 sections sont chargés (sur 24). L'analyse commence à l'appel de la fonction `main` et se termine à la ligne 8 avec une adresse de fin. Cette adresse de fin est produite par le script shell pour attraper la fin de la fonction `main`.

```

1 load sections .plt, .text, .rodata, .data, .got, .got.plt, .bss from file
2
3 secret global r, cin, y, x
4
5 starting from <main>
6
7 with concrete stack pointer
8 halt at 0x00000000000000464
9 explore all
10

```

Code 7 – Instructions permettant d'analyser le code 3 compilé vers RiscV-32

Ce modèle, qui nous servira de base pour la suite du développement, a permis une analyse rapide entre différents compilateurs et différentes architectures.

Application et observation entre architectures et compilateurs

FIGURE 3.2 – Tableau de résultats d'analyse Binsec pour architecture ARMv7 et ARMv8

opt\fonction analysée	cmovznz4				
Clang+LLVM	14.0.6	15.0.6	16.0.4	17.0.6	18.1.8
-O0	✓	✓	✓	✓	✓
-O1	✓	✓	✓	✓	✓
-O2	✓	✓	✓	✓	✓
-O3	✓	✓	✓	✓	✓
-Os	✓	✓	✓	✓	✓
-Oz	✓	✓	✓	✓	✓

✓ : binaire secure

On comprend, à la lecture du tableau 3.2, que la politique temps constant est considérée respectée par Binsec sur les versions testé ainsi que pour les différentes options de compilation. Ce résultat est encourageant pour la suite du projet.

FIGURE 3.3 – Tableau de résultats d'analyse Binsec pour architecture Risc-V

opt\fonction analysée	cmovznz4 - 64 bits		cmovznz4 - 32 bits	
Compilateur et architecture	gcc 15.1.0	clang 19.1.7	gcc 15.1.0	clang 19.1.7
-O0	~	×	~	×
-O1	✓	×	✓	×
-O2	✓	×	✓	×
-O3	✓	×	✓	×
-Os	✓	×	✓	×
-Oz	✓	×	✓	×

✓ : binaire secure; ~ : binaire unknown; × : binaire insecure

Les résultats dans le tableau 3.3 est indéniable : la version 19.1.7 de clang rend le code source perméable à des attaques temporelles.

Identification de défaut

Pour construire le tableau 3.3, plusieurs alertes se sont levées et on permis de mettre en évidence un bug présent dans Binsec. Cette erreur dans l'analyse symbolique provoquait l'arrêt de l'exploration par explosion de l'usage de la mémoire. Les registres `ld` (*load*) et `sd` (*store*) étaient mal gérés. En particulier l'opérande `ld`, simulé par un tableau, n'était jamais vidé. Cette découverte a amené un correctif et une amélioration de Binsec. De part l'envergure de ce projet, il est possible que d'autres erreurs dû à Binsec soient découvertes. L'exploration de nombreuses et nouvelles ISA^a, surtout avec Risc-V qui est encore en développement et perfectionnement, permet de renforcer cet outil plus efficacement et rapidement que par la construction de tests manuels.

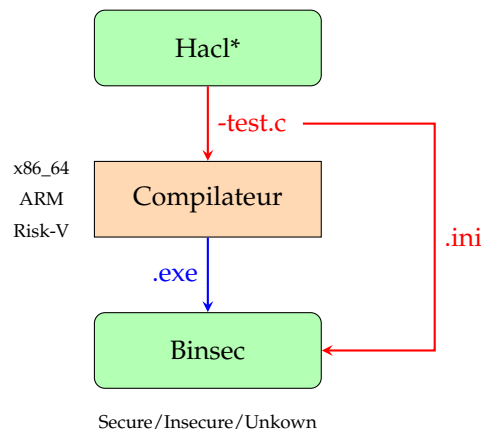
^a. Ancronyme anglais pour Architecture de Jeu d'Instruction, désigne l'ensemble des instructions assembleur associées à une architecture.

En explorant plus en avant le code binaire, on découvre que ces erreurs sont dus à l'opérande `beqz`⁴. L'ISA de Risc-V n'a pas à sa disposition un opérande comme `cmov` en X86_64 ou ARM. Donc l'application d'optimisation de compilation force l'usage de cette opérande qui n'est pas en temps constant. L'optimisation qui réalise ce changement se nomme «*Inst-CombinePass*».

On a pu observer ici une manifestation indéniable des précédents résultats proposés par d'autres travaux de recherche. Une solution serait de modifier l'ISA pour permettre cette opération d'être en temps constant. Celle qui a été retenue c'est d'employer un `pragma`, ici `# pragma clang optimise <off/on>`. Cette instruction donné dans le code source indique au compilateur de désactiver ses optimisation pour le code contenu entre les deux balise `off`, `on`. Cette solution entraîne des pertes de performance et des ralentissements quand au temps de compilation et à l'usage des ressources, il vaut mieux l'utiliser avec parcimonie que désactiver toutes les optimisations de compilations.

Nous avons notre besoin, nous avons nos exigences, nous avons fait des tests pour comprendre le processus que l'on va devoir automatiser. Ces idées peuvent être illustrées par la figure 3.4 : depuis Hacl*, on extrait une fonction que l'on veut tester en un fichier de test en C, on identifie les paramètres secrets et on conçoit le script pour Binsec, on compile notre fichier C à notre guise, on lance l'analyse Binsec.

FIGURE 3.4 – Flot de travail de l'outil d'analyse à concevoir



Transition

4. Effectue un branchement si la valeur du registre consulté est zéro. Cette opérande est propre à Risc-V.

Érysichthon à jamais affamé

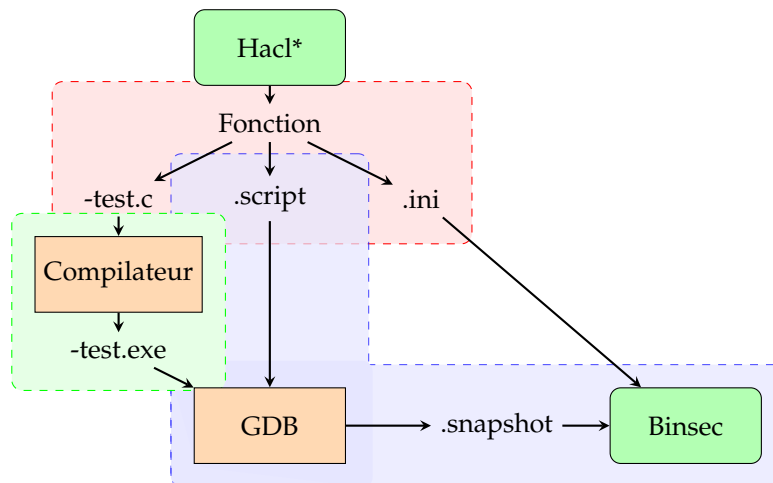
intro

4.1 Planification et préparations

Nous avons nos spécificités technique et nous savons qu'elle forme notre outil doit avoir. Nous pouvons commencer par synthétiser les opérations nécessaires.

Nous allons donc concevoir des protocoles pour identifier les étapes nécessaires pour que Binsec analyse entièrement un fichier et nous renvoie un parmi [secure, unknown, insecure]. Le protocole x86_64 est particulier. Depuis la version 0.5.0 de Binsec il est possible de fournir un «cliché mémoire»¹ pour accélérer l'analyse. On va se servir de cet avantage pour l'intégrer notre graphe d'exécution. La machine sur laquelle le projet sera développé est sur une architecture x86_64, cela nous permet d'utiliser l'outil GDB pour la génération de cliché mémoire.

FIGURE 4.1 – Protocole pour analyser des fichiers compiler en x86_64

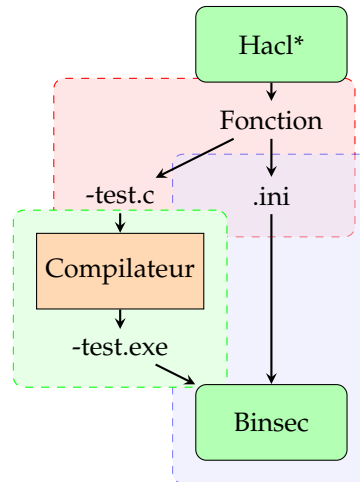


Ce graphe présente la chaîne d'étape à avoir pour obtenir une analyse Binsec depuis une fonction que l'on cible. Plusieurs zones sont distinguées. La zone verte correspond à l'étape de compilation, la zone bleue à l'étape de préparation de l'analyse et la zone rouge à la synthèse des fichiers de tests et d'instruction pour l'analyse. Ce choix de couleur est adapté à la difficulté attendu de chaque étape. L'opération de compilation consiste en une commande. L'opération de préparation d'analyse consiste aussi en deux commandes : un appel à GDB avec le binaire puis un appel à Binsec avec le cliché mémoire et les instruction d'analyse.

1. Plus couramment 'Core dump', terme technique anglais désignant une copie de la mémoire vive et des registres d'un programme. Ce fichier sert à être analysé, généralement par un débogueur.

Avec ce graphe réalisé, on peut le modifier pour préparer la voie à d'autres architectures. Dans un format plus générique voici comment se présente notre protocole d'analyse :

FIGURE 4.2 – Protocole générique d'analyse



Dans ce contexte, une question se pose : est-ce que la conception des script pour Binsec (.ini) est automatisable ou est-ce qu'il faudra utiliser des émulateurs pour générer des clichés mémoire et revenir dans le cas de la figure 4.1 ?

Cette question est importante parce que lorsqu'on analyse un fichier ARM par exemple, si le fichier source contient des appels à des fonctions systèmes, les `IFUNC`, celles-ci sont exécutées en fonction de l'architecture qui exécute le programme. Or Binsec n'a pas ces informations, il faut les fournir à la main. Voici le script nécessaire pour une vérification de la fonction «`Hacl_AEAD_Chacha20Poly1305_Simd128_encrypt`» compilé vers ARMv8.

```

load sections .plt, .text, .rodata, .data, .got, .got.plt, .bss from file

secret global input1, aad1

@[0x00000048f008,8] := <__memcpy_generic>
@[0x00000048f018, 8] := <__memset_generic>
@[0x00000048f030,8] := <__memcpy_thunderx2>

lr<64> := 0xdeadbeef as return_address

starting from <main>
with concrete stack pointer

halt at return_address
halt at <_IO_puts>
explore all

```

Code 8 – Script d'instruction pour analyser un binaire compilé vers ARM

Les lignes 5 à 7 sont présente pour indiquer les branchement à effectuer par Binsec lorsque qu'il rencontre ces adresses. Cette opération automatiquement exécutée lors de l'initialisation de l'exécution doit ici être précisée avec les fonction présente dans le binaire. Automatiser ces affectations peut être difficile et nécessiter quelques outils d'analyse supplémentaires pour attraper les adresses qui ont besoin d'être réaffecté et leur attribuer les fonctions les plus adaptées.

Nommer un outil

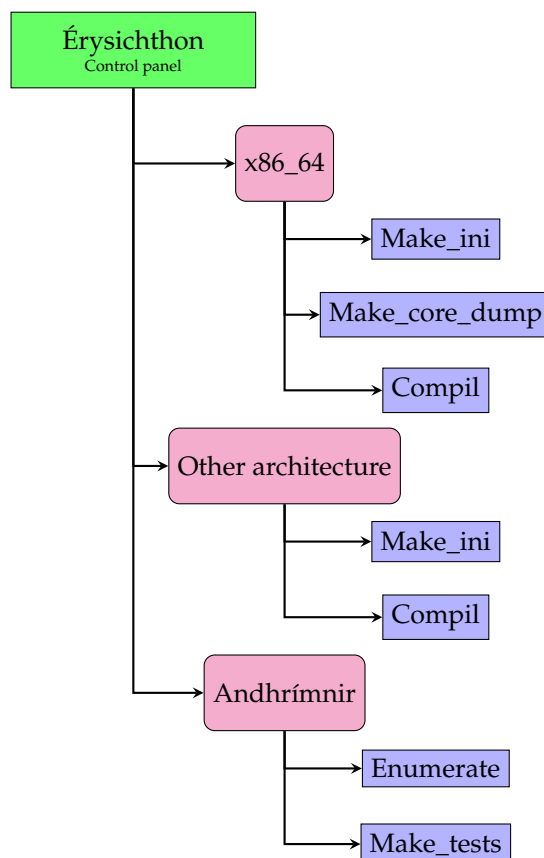
Rapidement il a fallu trouver un nom pour ce projet, l'appeler par "Notre outil..." devenait lourd et redondant entre les réunion hebdomadaire. En revanche trouver LE nom adéquat n'est pas une chose aisée, il peut être dû à une blague, une référence ou plus simplement liés au sens du projet. Dans notre cas, on aime la mythologie et le travail réalisé peut se résumer à "faut donner à manger à Binsec".

Érysichthon est une personnage de la mythologie grecque condamné à être affamé au point de se dévorer lui-même pour avoir détruit l'idole d'un dieu. Ce nom me plaît et sera retenu pour la suite du projet.

Conception d'Érysichthon

On a des protocoles. On a des expériences sur des fonctions de Hacl*. On a un nom. Il faut maintenant passer à l'échelle et concevoir un outil complet qui interagit et exécute nos instructions. L'outil sera en somme une combinaison de script python, shell et de Makefile. Il faut donc les organiser et adapter leurs interactions pour arriver au bout de nos peines.

FIGURE 4.3 – Structure des modules d'Érysichthon



On retrouve les différentes étapes de nos protocoles représentées par des modules au sein de leur noeuds respectifs : «Make_ini» pour les scripts pour Binsec, «Make_core_dump» pour la génération des clichés mémoires et «Compil» pour les appels aux compilateurs. Le dernier noeud «Andhrímnir» est particulier et détaillé dans la section suivante.

Andhrímnir

Ce module de Érysichthon est particulier. Premièrement c'est le seul qui est aussi son nom. Ce module consiste à produire les fichiers qui seront compilés puis analysés par Binsec. Ce module est nommé d'après le cuisinier de la mythologie nordique, car comme lui, il prépare sempiternellement des mets pour son maître.

Ce module est nommé car il constitue un projet dans le projet, sa seule conception pris plus de la moitié du temps de développement total. C'est un outil qui, à partir d'un pro-

jet en C, est capable de générer automatiquement des tests qui compilent et peuvent ensuite être proposés à des outils d'analyse binaire. Ce module est agrégé à Érysichthon mais peut être porté vers d'autres projets. À la différence des logiciels qui produisent des tests unitaires (uniquement sur des projets Java, Haskell ou C# et souvent associés à des offres payantes), on a ici une couverture complète des fonctions portés par le projet C.

Ce module, comme son grand frère, fonctionne et est abouti, par contre il nécessite quelques opérations manuelles et peut être amélioré pour pouvoir supporter d'autres projets C. Actuellement il possède quelques optimisations associées à Hacl* pour accélérer la réalisation de la preuve de concept.

4.2 Conception et usages

On va commencer par le petit frère, Andhrímnir. Il fonctionne avec une phase d'initialisation «Enumerate» et une phase de production de tests «Make_tests», elle-même découpée en plusieurs étapes. La génération de 548 fichiers de tests est réalisée en moins de 2 secondes.

Enumerate

Cette étape, de réalisation très simple, consiste à identifier toutes les fonctions qui auront un fichier test généré. Comme Hacl* génère automatiquement son code C, on peut exploiter cette particularité pour lister rapidement nos fonctions. L'opération actuellement réalisée est de lister l'ensemble des fichiers ".h" contenu dans le répertoire cible. Ensuite un parcours et une lecture de ceux-ci nous donne toutes les fonctions de l'API d'Hacl* et d'avoir une couverture complète du projet.

C'est lors de cette étape que l'on peut spécifier ou retirer des fichiers ou plus précisément des fonctions de la chaîne de production. Ce garde fou permet d'accélérer l'obtention des résultats finaux et d'aider grandement lorsque l'on souhaite déboguer.

Make_test

Le modèle de fichier de l'on veut construire est similaire aux tests minimaux que l'on a pu effectuer auparavant. Des paramètres, une déclaration de fonction et un appel à la fonction `exit`, c'est la recette pour une analyse simple. Binsec réalise une analyse symbolique, il ignore donc la valeur réelle des entrées. L'objectif est donc de concevoir un test qui compile pour avoir ensuite les bonnes instructions assembleurs à analyser. L'exemple 4.4 illustre comment sont générés nos fichiers de tests.

FIGURE 4.4 – Test de la fonction `Hacl_EC_K256_felem_sqr`

```
//
// Made by
// ANDHRÍMNIR - 0.3.0
// 09-07-2025
//
#include <stdlib.h>
#include "Hacl_EC_K256.h"

#define BUFFER_SIZE 5
uint64_t a[BUFFER_SIZE];
uint64_t out[BUFFER_SIZE];

int main (int argc, char *argv[]) {
    Hacl_EC_K256_felem_sqr(a, out);
    exit(0);
}
```

Phase introductive : 8 lignes

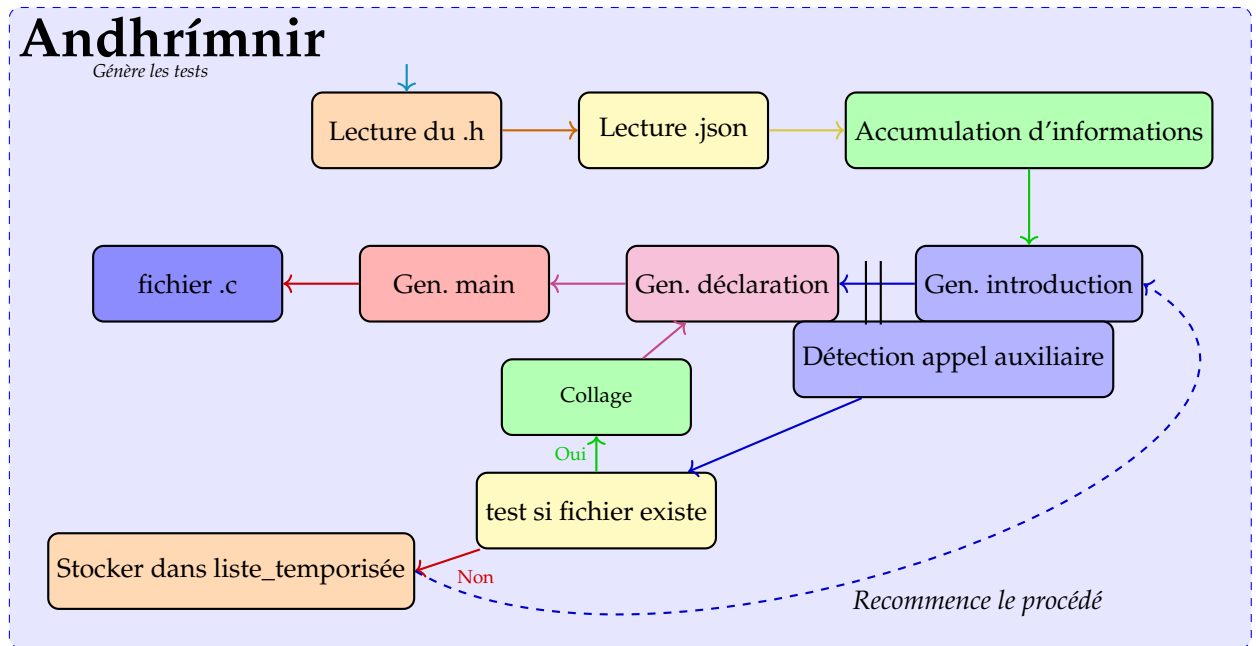
Phase déclarative

Phase principale

Une première partie initie la fichier. Cette partie contient les appels inclusifs de la librairie standard C, l'invocation de la librairie HACL* au travers du fichier d'en-tête de référence (ici `Hacl_EC_K256`) et la signature de fabrication en commentaires. L'utilisation de la librairie standard permet d'utiliser la fonction `exit`. Avec cet appel, on pourra construire nos scripts Binsec avec une interruption sur cette fonction. Cet arrêt précoce permet d'accélérer l'analyse du binaire de la cible (ici `Hacl_EC_K256_felem_sqr`) et nous garantir que cette analyse soit complète.

La deuxième partie contient tous les éléments déclaratifs nécessaire à l'invocation de la fonction. Puis on termine avec le corps du fichier C qui contient l'appel de la fonction, notre balise de fin avec la fonction `exit`. Cette construction est standardisé entre les fichiers et permet de mettre en places quelques optimisations.

FIGURE 4.5 – Schéma de conception d'Andhrímnir



Comme illustré par la figure 4.5, la génération des tests est effectuée lors de la lecture des fichiers d'en-tête. Une phase de lecture d'un fichier de données "json" ensuite est réalisée pour avoir toutes les informations nécessaires à la constitution du fichier test. Une fois ces étapes réalisées, Andhrímnir commence sa préparation. Or il se trouve que quelques fois, certaines fonctions de HACL* font appel à des structures propres à la librairie qui ont une instantiation particulière. Le module «Détection appel auxiliaire» permet de vérifier ce cas de figure.

Dans le cas où aucun appel n'est détecté, Andhrímnir continue sa préparation avec les étapes successives illustrées par la figure 4.4 : génération des déclarations puis du main.

À l'inverse où un appel est détecté, il est possible la fonction soit déclarée dans un autre fichier d'en-tête. Si c'est le cas, alors Andhrímnir doit déterminer quel fichier contient les informations requises pour compléter les informations nécessaires pour produire un fichier de test correct. La solution qui nous est venue est de temporiser le problème. Andhrímnir prépare des tests pour toutes les fonctions. Donc s'il a besoin d'une fonction qu'il a déjà préparé, on peut accéder aux informations contenues dans le fichier de test associé ; s'il a besoin d'une fonction qu'il n'a pas encore préparé, alors il peut la mettre de côté et retravailler dessus une fois qu'il a fini son premier passage sur toutes les fonctions d'HACL*. Ce procédé est récursif pour pallier le problème d'appels en cascade.

En réalité Andhrímnir ne va pas recharger les informations d'une fonction qu'il a besoin, il va faire cette opération de «collage». Elle consiste à une instruction shell qui vient ajouter (coller) au fichier en cours de conception la partie déclaration du fichier. Cette astuce permet d'éviter une nouvelle étape d'accumulation d'informations.

La phase de lecture dans les fichiers données json existe pour accélérer le développement de Andhrímnir. Cela permet de venir ajouter manuellement des instructions haut niveau pour la conception des tests. Le code en annexe 9 illustre ce point. Une fonction a besoin de que les éléments qui lui sont données en paramètres respecte ses attentes. Cet exemple est accompagnée du fichier json associé et du fichier de tests final ??.

Make_core_dump et Compilation

Les opérations de production de clichés mémoire et de compilation sont un assemblage de commandes shell et script pour GDB qui sont concevable sans problèmes. L'élément difficile à cet étape est la compilation de la bibliothèque Hacl*. Cette étape est nécessaire pour correctement compiler nos fichiers tests qui appellent Hacl*. Or cette gestion de la compilation est réalisée par le projet Hacl* lui-même et à besoin d'être amélioré pour permettre une compilation croisée vers d'autres architectures.

Une modification du script de compilation «*configure*» a été proposé et modifié sur le dépôt officiel du projet Hacl*.

Make_ini

Ce module consiste à concevoir les fichiers d'instructions pour Binsec. Il doit spécifier les variables secrètes associé à la fonction analysé. À la suite des exemples cités précédemment, le code 12 illustre comment

Références

TABLE 5.1 – Liste des options de compilations et leurs effets (non exhaustive), <https://gcc.gnu.org/onlinedocs/gcc/Optimize-Options.html>

Option de compilation	Effet
-O0	Compile le plus vite possible
-O1 / -O	Compile en optimisant la taille et le temps d'exécution
-O2	Comme -O1 mais en plus fort, temps de compilation plus élevé mais exécution plus rapide
-O3	Comme -O2, avec encore plus d'options, optimisation du binaire
-Os	Comme -O2 avec des options en plus, réduction de la taille du binaire au détriment du temps d'exécution
-Ofast	optimisations de la vitesse de compilation
-Oz	optimisation agressive sur la taille du binaire

Conception de tests

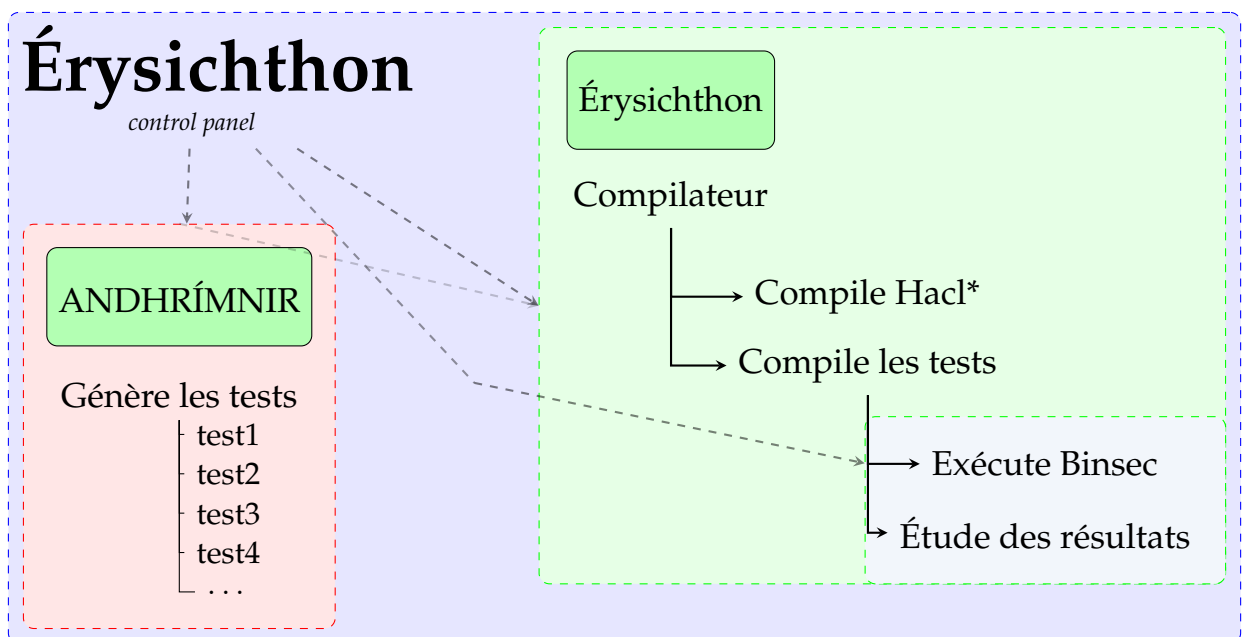


FIGURE 6.1 – Structure d'Érysichthon, schéma du point de vue de l'utilisateur
Les flèches grises indiquent tous les éléments actionnables individuellement.

```

1  /**
2  Encrypt a message `input` with key `key`.
3
4  The arguments `key`, `nonce`, `data`, and `data_len` are same in encryption/decryption.
5  Note: Encryption and decryption can be executed in-place, i.e.,
6  `input` and `output` can point to the same memory.
7
8  @param output Pointer to `input_len` bytes of memory where the ciphertext is written to.
9  @param tag Pointer to 16 bytes of memory where the mac is written to.
10 @param input Pointer to `input_len` bytes of memory where the message is read from.
11 @param input_len Length of the message.
12 @param data Pointer to `data_len` bytes of memory where the associated data is read from.
13 @param data_len Length of the associated data.
14 @param key Pointer to 32 bytes of memory where the AEAD key is read from.
15 @param nonce Pointer to 12 bytes of memory where the AEAD nonce is read from.
16 */
17 void
18 Hacl_AEAD_Chacha20Poly1305_Simd256_encrypt (
19     uint8_t *output,
20     uint8_t *tag,
21     uint8_t *input,
22     uint32_t input_len,
23     uint8_t *data,
24     uint32_t data_len,
25     uint8_t *key,
26     uint8_t *nonce
27 );

```

Code 9 – Déclaration de la fonction **encrypt** dans le fichier d'en-tête Hacl_AEAD_Chacha20Poly1305_Simd256.h

```

1  {
2  "Meta_data": {
3      "build" : "13-06-2025",
4      "version" : "0.2.0"
5  }
6
7  , "Hacl_AEAD_Chacha20Poly1305_Simd128_encrypt": {
8      "*output": "BUF_SIZE"
9      , "*input": "BUF_SIZE"
10     , "input_len": "BUF_SIZE"
11     , "*data": "AAD_SIZE"
12     , "data_len": "AAD_SIZE"
13     , "*key": "KEY_SIZE"
14     , "*nonce": "NONCE_SIZE"
15     , "*tag": "TAG_SIZE"
16     , "BUF_SIZE": 16384
17     , "TAG_SIZE": 16
18     , "AAD_SIZE": 12
19     , "KEY_SIZE": 32
20     , "NONCE_SIZE": 12
21   }
22 }

```

Code 10 – Extrait du fichier Hacl_AEAD_Chacha20Poly1305_Simd256.json

```

1  //
2  // Made by
3  // ANDHRÍMNIR - 0.5.4
4  // 12-08-2025
5  //
6
7  #include <stdlib.h>
8  #include "Hacl_AEAD_Chacha20Poly1305_Simd128.h"
9
10 #define BUF_SIZE 16384
11 #define TAG_SIZE 16
12 #define AAD_SIZE 12
13 #define NONCE_SIZE 12
14 #define KEY_SIZE 32
15 uint8_t output[BUF_SIZE];
16 uint8_t tag[TAG_SIZE];
17 uint8_t input[BUF_SIZE];
18 uint32_t input_len_encrypt = BUF_SIZE;
19 uint8_t data[AAD_SIZE];
20 uint32_t data_len_encrypt = AAD_SIZE;
21 uint8_t key[KEY_SIZE];
22 uint8_t nonce[NONCE_SIZE];
23
24
25 int main (int argc, char *argv[]){
26   Hacl_AEAD_Chacha20Poly1305_Simd128_encrypt(output, tag, input, input_len_encrypt,
27     data, data_len_encrypt, key, nonce);
28     exit(0);
29 }

```

Code 11 – Code généré du fichier test Hacl_AEAD_Chacha20Poly1305_Simd256_encrypt.c

```

1  starting from core
2
3  secret global output, input, data, key, nonce, tag
4  replace opcode 0f 01 d6 by
5  zf := true
6  end
7  replace opcode 0f 05 by
8  if rax = 231 then
9    print ascii "exit_group"
10   print dec rdi
11   halt
12 end
13 print ascii "syscall"
14 print dec rax
15 assert false
16 end
17 halt at <exit>

```

Code 12 – Instruction Binsec générée automatiquement,
Hacl_AEAD_Chacha20Poly1305_Simd256_encrypt.ini