

# **Opake Prädikate gegen symbolische Ausführung mittels stochastischer Unentscheidbarkeit**

Teilnehmende:	Paul Baumgartner (18 J.)
Erarbeitungsort:	Hildesheim
Projektbetreuende:	Dr. Arndt Latußeck
Fachgebiet:	Mathematik/Informatik
Wettbewerbssparte:	Jugend forscht
Bundesland:	Niedersachsen
Wettbewerbsjahr:	2026

# Inhaltsverzeichnis

<b>1</b>	<b>Projektüberblick</b>	<b>0</b>
<b>2</b>	<b>Einleitung</b>	<b>1</b>
<b>3</b>	<b>Theoretische Grundlagen</b>	<b>1</b>
3.1	Obfuskation . . . . .	1
3.2	Opake Prädikate . . . . .	1
3.3	Symbolische Ausführung . . . . .	2
3.4	Kumulative Verteilungsfunktionen . . . . .	3
<b>4</b>	<b>Hintergrund und Motivation</b>	<b>3</b>
<b>5</b>	<b>Ansatz</b>	<b>4</b>
5.1	Angreifermodell . . . . .	4
5.2	Stochastische opake Prädikate . . . . .	4
5.3	Algorithmus . . . . .	5
<b>6</b>	<b>Implementierung</b>	<b>6</b>
6.1	Entscheidungen . . . . .	6
6.2	Generierung von Pseudozufallsvariablen . . . . .	7
6.3	Wahrscheinlichkeitsverteilungsgenerierung . . . . .	7
6.4	Generierung von ununterscheidbarem Füllcode . . . . .	8
<b>7</b>	<b>Evaluierung</b>	<b>8</b>
7.1	Kriterien . . . . .	8
7.2	Vergleich mit existierenden Obfuskationsmethoden . . . . .	8
<b>8</b>	<b>Ergebnisdiskussion</b>	<b>8</b>
<b>9</b>	<b>Fazit und Ausblick</b>	<b>8</b>
	<b>Literaturverzeichnis</b>	<b>B</b>
	<b>Abbildungsverzeichnis</b>	<b>B</b>

# **1 Projektüberblick**

In diesem Projekt wird sich mit der Softwareobfuskation, also dem Verschleiern/Verstecken wichtiger Funktionalitäten von Programmen, beschäftigt. Konkret wird eine neue Art der Obfuskation vorgestellt, welche sich als resistent gegenüber existierender Methoden zur Deobfuskation erweist. Hierfür werden Wahrscheinlichkeiten in Bedingungen innerhalb des Programmes verknüpft, wobei für einen Angreifer, welcher das Programm verstehen möchte, unklar ist, inwiefern diese Bedingungen wahr oder falsch sind. Obwohl die Erfüllung dieser Bedingungen zufällig ist, werden Wahrscheinlichkeiten verwendet, welche in der Praxis garantieren, dass immer das gewünschte Verhalten auftritt.

## 2 Einleitung

*Obfuskation* (lat. *obfuscare*: verdunkeln) bezeichnet jede Transformation von Programmen zur Hinderung von sog. Reverse Engineering - der Analyse von Software zum Cracken, Verstehen oder Kopieren. Obfuskation kommt zum Einsatz in der Malwareentwicklung - um vor Detektion von sog. EDRS zu schützen, in der Industrie - um vor Kopien von Softwarefunktionen sowie vor Cracking zu schützen und im Militär - um dem Feind ein Verständnis der eigenen Waffensysteme zu behindern. Da das Programm hierbei noch die Ursprüngliche Semantik beibehält kann jede Software mit genügend Zeit, Aufwand und Geld trotz Obfuskation verstanden werden. Der Sinn von Obfuskation ist also nicht die komplette Verhinderung von *Reverse Engineering*, sondern vielmehr dieses wirtschaftlich unrentabel zu machen.

Diese Arbeit fokussiert sich auf eine Art der Obfuskation, der Kontrollflussobfuskation, und der meist verbreiteten Angriffsweise, um diese zu bekämpfen. ...

Diese Arbeit nimmt ein mathematisch-informatisches Grundwissen an Assembler, Kompilern, grundlegender Zahlentheorie und linearer Algebra an. Matrizen und Vektoren werden fettgedruckt, wobei Matrizen großgeschrieben werden (z.B. **A**) und Vektoren klein (z.B. **b**).

## 3 Theoretische Grundlagen

### 3.1 Obfuskation

Collberg et al. [2] definieren Obfuskation wie folgt:

**Definition 3.1** (Obfuskation). Sei  $P \xrightarrow{\mathcal{T}} P'$  eine Transformation eines *Quellprogrammes*  $P$  zu einem *Zielfprogramm*  $P'$ . Eine solche Transformation ist eine Obfuskation, wenn das obfuskierete Programm  $P'$  dasselbe beobachtbare Verhalten wie  $P$  für den Endnutzer aufweist.

Per Definition sind somit Nebenwirkungen (z.B. Herunterladen von neuen Daten etc.) erlaubt, solange sie nicht vom Nutzer erfahren werden. Die präsentierte Methode dieser Publikation nutzt diese Lockerung der Einschränkungen auf obfuskierende Transformationen stark aus, wie später ersichtlich sein wird.

Um eine Obfuskationsmethode zu evaluieren werden typischerweise die Metriken *Stärke*, *Resilienz*, *Kosten* und *Tarnung* verwendet [2].

### 3.2 Opake Prädikate

Die folgenden Definitionen sind aus [9] sowie vom Pionierwerk [3] modifiziert übernommen. Es wird sich auf hierbei auf invariante opake Prädikate beschränkt. Ferner wird angenommen, dass alle Funktionen/Prädikaterme evaluierbar sind.

**Definition 3.2** (Opake Prädikate). Sei  $O : \Phi \rightarrow \{0, 1\}$  eine Abbildung einer Variable  $\phi \in \Phi$  zu einem Prädikat. Das Prädikat  $O(\phi)$  ist opak, wenn für alle  $\phi \in \Phi$  gilt, dass  $O(\phi)$  denselben Wert (1 oder 0 bzw. wahr oder falsch) hat.

In anderen Worten: Das Prädikat  $O(\phi)$  ist opak, wenn dessen Wert für alle möglichen Parameter *a priori* bestimmt ist (also für den Programmierer bekannt ist) aber für ein Verständnis einer weiteren Person (ein Angreifer) *a posteriori* (durch Beobachtung) zu bestimmen ist.

Diese Arbeit unterscheidet zwischen zwei Arten opaker Prädikate:

**Definition 3.3.** Sei  $O : \Phi \rightarrow \{0, 1\}$  ein opakes Prädikat.

$O(\phi)$  ist vom Typ

1.  $P^T$ , wenn für alle  $\phi \in \Phi$  gilt:  $O(\phi) = 1$  bzw. *wahr*.
2.  $P^F$ , wenn für alle  $\phi \in \Phi$  gilt:  $O(\phi) = 0$  bzw. *falsch*.



Abbildung 1: Kontrollflussgraph einer einfachen Funktion mit opakem Prädikat. Abbildung aus der Disassembly des Spiels *Overwatch* mittels IDA entnommen.

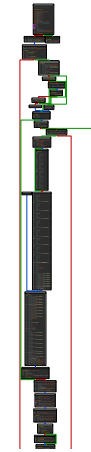


Abbildung 2: Ausschnitt des Kontrollflussgraphen einer Funktion mit vielen opaken Prädikaten. Abbildung aus der Disassembly des Spiels *Overwatch* mittels IDA entnommen.

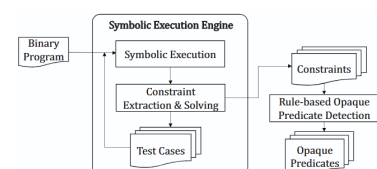
Opake Prädikate werden in der Softwareobfuskation eingesetzt, um ein Verständnis über den Kontrollfluss des Programms zu behindern. Damit opake Prädikate als Obfuskationsmethode<sup>1</sup> genutzt werden können, müssen sie wiederholt angewandt werden. Dadurch entsteht ein komplexerer Kontrollflussgraph und der Angreifer weiß folglich nicht, welche Basisblöcke zu analysieren sind. Die Stärke der opaken Prädikate ist hierbei abhängig von der Stärke ihres Terms/Ausdrucks. Mit zunehmender Komplexität der Prädikate und zunehmender Anzahl dieser, nimmt also auch die Obfuskationsstärke (Verwirrung und Unverständnis) beim Angreifer zu.

**Beispiel 1.** Das Prädikat  $O(\phi) = (-1977224191 \& 1 == 1)$  aus Abb. 1, wobei "&" dem bitweisen "und" Operator entspricht, ist sehr einfach. Eine Berechnung genügt, um zu erkennen, dass das Prädikat immer wahr ist.

**Beispiel 2.** Das Prädikat  $O(\phi) = (y < 10) \vee (x \cdot (x + 1) \bmod 2 == 0)$  aus [4] mit  $\phi = (x, y)$  und  $x, y \in \mathbb{Z}$  ist etwas komplexer. Auch dieses Prädikat ist immer wahr, da  $x \cdot (x + 1)$  immer gerade ist. Der Wert von  $y$  ist folglich egal.

### 3.3 Symbolische Ausführung

Nach dem aktuellen Stand der Forschung scheint symbolische Ausführung der vielversprechendste Ansatz für die Bekämpfung opaker Prädikate zu sein. Ein Symbolischer Ausführungseengine besteht aus 2 Hauptkomponenten: einem zentralen symbolischen Ausführungsmodul und einem Modul zur Lösung von Einschränkungen (*constraint solver*).



<sup>1</sup>D.h., dass der wirkliche Pfad, welcher von einem opaken Prädikat verschleiert wird, nicht einfach erkannt werden kann

**Abbildung 3:** Konzeptuelles Framework zur Erkennung opaker Prädikate mit symbolischer Ausführung, Abb. aus [10] übernommen

Es gibt zwei Arten symbolischer Ausführungsmodule: statisch und dynamisch (letzteres wird auch concolische symbolische Ausführung genannt). Concolische Ausführungsengines wie BAP und Triton führen zunächst das Programm mit konkreten Werten aus und führen dann eine symbolische Analyse der generierten Befehlsspuren durch. Statische symbolische Ausführungsengines "heben" zuerst die Assembly-Anweisungen des Programmes in eine abstraktere Zwischensprache ("intermediate representation") an und führt dann eine symbolische Ausführung dessen mit statischen Analyseansätzen durch. Diese Methode wird erfolgreich in Angr verwendet.

### 3.4 Kumulative Verteilungsfunktionen

**Definition 3.4** (Kumulative Verteilungsfunktion). Eine Funktion  $F$  ist eine kumulative Verteilungsfunktion, wenn gilt:

- (1)  $F$  ist streng monoton steigend:
- (2)  $F$  ist rechtsstetig:
- (3)  $\lim_{x \rightarrow -\infty} F(x) = 0$  und  $\lim_{x \rightarrow +\infty} F(x) = 1$

## 4 Hintergrund und Motivation

Der folgende Abschnitt soll einen Überblick über existierende Arten opaker Prädikate liefern und diese in Bezug auf ihre Qualität evaluieren. Hierfür werden die Faktoren aus Abschnitt 3.1 verwendet.

	Stärke	Resilienz	Kosten	Tarnung
Obfuscator-LLVM [4]	X	trivial	sehr gering	hoch
<i>Bi-Opake</i> Prädikate [10]	X	mittel	hoch	hoch
<i>Exceptions</i> [6]	X	mittel	sehr hoch	gering
NP-schwere Probleme [ <b>&lt;empty citation&gt;</b> ]	X	gering	mittel	mittel
ungelöste Probleme [7]	X	hoch	sehr hoch	hoch

Tabelle 1: Vergleich existierende Arten opaker Prädikate.

Hier objektiver werden, evtl. "Literature Reviews" zitieren und zusammenfassen

- Obfuscator-LLVM ✓
- Bi-Opaque Prädikate ✓
- Linear Obfuscation to Combat Symbolic Execution
- Binary Code Side Effects ✓
- "When Are Opaque Predicates Useful"
- "Software obfuscation on a theoretical basis and its implementation," IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, 2003.
- (Collberg et al.) Tigress (dynamic opaque prediactes!!)

## 5 Ansatz

### 5.1 Angreifermodell

Diese Arbeit geht aufgrund der Ähnlichkeit behandelter Thematik von einem ähnlichen Angreifermodell wie [10] aus. Ein Angreifer hat direkten Zugriff auf das Programm und dessen Anweisungen. Es sei dem Angreifer hierbei nicht vorgegeben, wo und inwiefern das Programm obfuskiert ist. Der Angreifer kann das Programm nur statisch mittels symbolischer Ausführung sowie durch Untersuchung der Programmanweisungen analysieren. Pattern Matching ist daher möglich. Der Angreifer kann das Programm nicht ausführen. Solchen Situationen begegnet man z.B. bei Malware, welche gegen Virtuelle Maschinen ("virtual machines") gehärtet ist oder bei Software, welche für proprietäre und unverfügbare Systeme geschrieben ist. Eine Härtung gegen weitere (dynamische) Analysemethoden des Angreifers wird in Abschnitt 9 diskutiert.

### 5.2 Stochastische opake Prädikate

Während andere Methoden wie [10] auf den praktischen Implementierungslimitationen symbolischer Ausführungseines basieren, nutzt diese Methode deren fundamentale Grenzen aus.

Die symbolischen Ausführungseines zugrundeliegenden SMT-Solver können im Wesentlichen folgendes:

1. Den Wahrheitswert einer Aussage herauszufinden.
2. Variablen mögliche Werte zuzuweisen, damit diese Aussage erfüllt ist.

In der Regel genügt dies für die Untersuchung deterministischer Algorithmen und entscheidbarer Probleme. Um den Wahrheitswert einer Aussage zu finden, muss sie für jeden möglichen Wert jeder Variable vom symbolische Ausführungseine bewiesen werden.

Will man allerdings Aussagen über einen probabilistischen Algorithmus treffen, so ist dies ohne erhebliche manuelle Eingriffe eines Nutzers in Form von extra Beschränkungen (*constraints* unmöglich. Jede zufällige Variable eines jeden Schleifenaufwurfes muss vom constraint solver als symbolische Variable betrachtet und somit für alle möglichen Werte überprüft werden. Bei einem Monte-Carlo-Algorithmus heißt dies allerdings, dass auch unwahrscheinliche Variablenwerte, welche zu *falschen* Ergebnissen führen, überprüft werden. Der Wahrheitsgehalt wird somit zwar formell *logisch* korrekt bewiesen - praktisch allerdings nicht.

**Beispiel 3.** *Man betrachte einen Algorithmus, welcher mit einer Wahrscheinlichkeit von 99,9999999999% den Wert 1 zurückgibt und mit einer Wahrscheinlichkeit von 0,0000000001 den Wert 0 zurückgibt.*

---

**Algorithm 1** Beispiel eines probabilistischen Algorithmus

---

```
1: procedure Foo()
2:    $x \leftarrow \text{UNIFORMRAND}(0, 1)$ 
3:   if  $x \leq 0,999999999999$  then
4:     return 1
5:   end if
6:   return 0
7: end procedure
```

---

*Nutzt man einen symbolischen Ausführungseengine, um zu prüfen, ob dieser Algorithmus den Wert 1 wiedergibt, so würde dieser behaupten, dass dies falsch sei.*

Dies bildet die Grundidee stochastischer opaker Prädikate. Anstatt Prädikate zu bilden, welche für alle Werte ihrer Parameter *wahr* bzw. *falsch* sind, werden Prädikate erzeugt, deren gewünschter Wert so wahrscheinlich ist, dass das Gegenteil praktisch nie auftritt.

**Definition 5.1** (Stochastische opake Prädikate). Sei  $O : \Phi \rightarrow \{0, 1\}$  eine Abbildung einer Variable  $\phi \in \Phi$  zu einem Prädikat. Das Prädikat  $O(\phi)$  ist stochastisch opak, wenn der Fall  $A \in \{0, 1\}$  mit einer so hohen Wahrscheinlichkeit eintritt, dass der Fall  $\bar{A}$  vernachlässigbar ist.

**Definition 5.2.**  $\mathcal{P}_{stoch}$  ist die Klasse aller stochastischen opaken Prädikate.

Zu  $\mathcal{P}_{stoch}$  gehören gemäß Definition 5.2 also sowohl Prädikate wie  $\text{UNIFORMRAND}(0, 1) \leq 0,999999999999 == 1$  aus Algorithmus 1 als auch Prädikate wie  $\text{POIS}(\lambda = 5) \geq 10 == 0$ .

Um zu garantieren, dass das Programm, in welchem das stochastische opake Prädikat eingefügt wurde, weiterhin funktioniert, kann für den ungewünschten Gegenfall praktische eine Wahrscheinlichkeit eingesetzt werden, welche unter der eines Hardwarefehlers liegt. Auch gewöhnliche Programme bzw. Computer können spontan versagen. Durch das Nutzen so geringer Wahrscheinlichkeiten ...

### 5.3 Algorithmus

Für jedes zu generierende opakes Prädikat wird im Programm ein Pseudozufallsvariable  $x$  generiert. Verschiedene Methoden hierfür werden in Abschnitt 6.2 gegeben. Wichtig ist, dass der Angreifer den Wert von  $x$  nicht statisch bestimmen kann. Es wird angenommen, dass  $x$  gleichverteilt ist. Diese Verteilung wird nun in eine andere (z.B. Normal-, Exponential-, Bernoulliverteilung) transformiert. Über diese können nun wahrscheinliche bzw. unwahrscheinliche Aussagen getroffen werden. Abschnitt 5.3 zeigt die

---

#### Algorithm 2 Beispiel eines stochastisch opaken Prädikats

---

**Require:**  $x$ : zufällige, unbekannte Zahl

```

1: SkalierteZufallszahl  $\leftarrow$  ABBILDUNG_AUF_EINHEITSINTERVALL( $x$ )           ▶ Einheitsintervall: [0; 1]
2: PoissonVerteilt  $\leftarrow$  INVERSECDF(SkalierteZufallszahl,  $\lambda = 5$ )
3: if PoissonVerteilt  $\leq 10$  then
4:     ...                               ▶ Praktisch immer wahr ( $P(X \leq 10) = 0,9863$ )
5: else
6:     ...                               ▶ Praktisch nie wahr ( $P(X \geq 10) = 0,0137$ )
7: end if
```

---

Umsetzung eines stochastischen opaken Prädikats in einem Programm.

Der Pseudocode für die Generierung solcher stochastischer opaker Prädikate ist in Algorithmus 3 beschreiben.



---

**Algorithm 3** Generierung stochastischer opaker Prädikate

---

```
1: Erstelle Funktion TransformiereVerteilung(x) im Modul, welche die gleichverteilte Variable  $x \in ]0; 1]$ 
...
2: procedure FÜGEPRÄDIKATEIN(Modul, T)
3:   for all Funktion in Modul do
4:     BasisBlock  $\leftarrow$  ZUFÄLLIGERBASISBLOCK(Funktion)
5:     Entry  $\leftarrow$  ZUFÄLLIGEANWEISUNG(BasisBlock)
6:     ZufaelligerWert  $\leftarrow$  ERSTELLEFUNKTIONSAUFRUF(Gleichverteilung, ]0; 1])
7:
8:     TransformierteVariable  $\leftarrow$  ERSTELLEFUNKTIONSAUFRUF(TransformiereVerteilung, ZufaelligerWert)
9:     ERSTELLEVERGLEICH(TransformierteVariable)
10:
11:     BBimmer_wahr  $\leftarrow$  SPLITBLOCK(BasisBlock, Entry)           ▶ Basisblöcke für Prädikatenfälle
generieren
12:     BBnie_wahr  $\leftarrow$  NEUERBLOCK(Funktion)
13:
14:     ERSTELLEBEDINGTEVERZWEIGUNG("TransformierteVariable  $\geq T$ ", BBimmer_wahr, BBnie_wahr)
15:
16:     Fülle Basisblock BBnie_wahr mit zufälligen Anweisungen.
17:   end for
18: end procedure
```

---

## 6 Implementierung

### 6.1 Entscheidungen

Zur Implementierung kamen drei Ansätze in Betracht: die Entwicklung eines Bin2Bin-Obfuskators<sup>2</sup>, die Implementierung in Form eines LLVM-Passes [5, 8]<sup>3</sup> sowie die Quellcodemanipulation<sup>4</sup>. Es wurde eine Entscheidung für einen LLVM-Pass getroffen aufgrund folgender Vorteile:

Durch die Nutzung von LLVM profitiert man von Abstraktion. Anstatt manuell Assembler-Anweisungen hinzuzufügen, Relocations zu managen und Support für verschiedene Betriebssysteme und Prozessorarchitekturen manuell zu übernehmen, kann dies indirekt einer Gruppe kompetenter Experten übergeben werden, welche sich aktiv damit befassen. Auch wenn andere Methoden wie ein Bin2Bin-Obfuskator direktere Kontrolle bieten würden, ist es fraglich ob eine Person in relativ kurzer Zeit verschieden gleiche Funktionalitäten besser implementieren könnte.

Ähnliches gilt für die Effizienz kompilierter Programme: Durch die abstrakte Zwischensprache in LLVM profitiert man von der konstanten Verbesserung tausender LLVM-Contributer, welche dafür sorgen, dass die eigenen Programme nahezu optimal kompiliert werden. Auch das ledigliche Einfügen vieler Prädikate an zufälligen Stellen kann in größeren Programmen bereits zu erhöhten Laufzeitkosten führen. In Situationen wie der Entwicklung für Embedded Systems, IoT und Echtzeitsystemen, können leicht erhöhte Laufzeiten gegen die Anwendung einer Obfuskation sprechen. Solchen Situationen begegnet man z.B. wie bei präziser optimierter medizinischer Technik sowie im Finanzwesen beim quantitativen Trading. In beiden Situationen kann eine leichte Verzögerung fatal sein - für Unternehmen und Personen.

Obfuskierter Programme werden also effizienter und ihre Implementierung schneller, da sich auf die

---

<sup>2</sup>Direkte Manipulation von Assembler-Anweisungen existierender Programme.

<sup>3</sup>Nutzung des LLVM-Projekts, um einen sog. *Compiler-Pass* zu schreiben.

<sup>4</sup>durch z.B. C-Makros und das Einfügen von inline Assembler-Ausschnitten

Obfuskation selbst fokussiert werden kann.

## 6.2 Generierung von Pseudozufallsvariablen

Damit der vorgestellte Ansatz funktionieren kann, bedarf er einer (Pseudo-)Zufallszahl, welche auf dem Einheitsintervall liegt und zugleich als unbestimmte symbolische Variable von symbolischen Ausführungseingines betrachtet wird. Hierfür werden die Parameter der obfuskierten Funktionen verwendet. Die Idee hierhinter ist, dass um diese zu bestimmen, ...Hat die zu obfuskierte Funktion kein Parameter, so kann ein Pseudoparameter hinzugefügt werden. Um zu garantieren, dass die generierte Pseudozufallszahl im Einheitsintervall liegt, existieren mehrere Methoden. Es wurde sich für die einfachste entschieden. Hier wird der erste Parameter als 64-Bit Gleitkommazahl betrachtet und durch dessen maximalen Wert geteilt.

## 6.3 Wahrscheinlichkeitsverteilungsgenerierung

Um den Ansatz gegen Pattern Matching resistent zu machen, wird er generalisiert. Anstatt sich auf eine Wahrscheinlichkeitsdichtefunktion bzw. Umkehrfunktion der kumulativen Verteilungsfunktion zu beschränken, soll für jedes stochastisches opakes Prädikat eine neue Wahrscheinlichkeitsverteilung verwendet werden. Mehrere Methoden kommen hierfür infrage:

**Generierung über Verteilungsklasse** lorem ipsum dolor sit amet... Der Nachteil hiervon ist, dass ...

Eine Alternative ist die eigenständige Generierung einer zufälligen Funktion  $F$ , welche die Eigenschaften einer kumulativen Verteilungsfunktion (vgl. Definition 3.4) erfüllt. Das allgemeine Vorgehen hierfür ist Folgendes:

1. Wähle ein zufälliges Intervall  $[a; b]$ .
2. Teile  $[a; b]$  in  $n$  Teile ein.
3. Lege  $F(a) = 0$  und  $F(b) = 1$  fest.
4. Für jede Stelle  $x_i \in [a; b]$  der  $n$  Stellen: wähle einen Punkt  $y_i$  für  $F$  an dieser Stelle, sodass  $y_i \in [0; 1]$  maximal  $\frac{1}{n}$  über  $y_{i-1}$  liegt.
5. Führe eine Regression für die Punkte durch.

Es genügt hierbei nicht, eine einfache Polynomregression durchzuführen, da hier Oszillationen zwischen den ausgewählten Punkten  $(x_i, y_i)$  entstehen können. Hierdurch wäre Bedingung 1 (vgl. Definition 3.4) nicht erfüllt. Um dieses Problem zu lösen, gibt es im Wesentlichen 2 Methoden:

1. **Splines:** Regression so durchführen, dass  $F$  stückweise definiert ist über mehrere kleinere Funktionen, welche alle notwendigen Bedingungen erfüllen. Hierfür gibt es mehrere mögliche Regressionsarten.
2. **Bernstein-Polynom-Regression:** Auf dem skalierten Intervall  $[0, 1]$  parametrisieren wir  $F$  durch ein Bernstein-Polynom

$$B_n(x) = \sum_{k=0}^n c_k \binom{n}{k} x^k (1-x)^{n-k}.$$

- (1) Monotonie von  $B_n$  folgt, wenn die Koeffizienten  $c_k$  nicht fallend sind (d. h.  $c_0 \leq c_1 \leq \dots \leq c_n$ ).
- (2) Stetigkeit ist gegeben, da es sich um eine Funktion handelt.
- (3) Da  $F(a) = 0$  und  $F(b) = 1$  ist und der Definitionsbereich von  $F$  das Intervall  $[0; 1]$  ist, ist  $\lim_{x \rightarrow -\infty} F(x) = 0$  und  $\lim_{x \rightarrow +\infty} F(x) = 1$ .

Das Fitten <sup>5</sup> kann als *Non-Negative Least-Squares*-Optimierungsproblem formuliert werden:

$$\begin{aligned} \min_{c \in \mathbb{R}^{n+1}} \|\mathbf{A}c - \mathbf{y}\|_2^2 \\ \text{u. d. N. } \mathbf{D}c \geq 0, \quad 0 \leq c_k \leq 1, \end{aligned} \tag{1}$$

wobei die Designmatrix  $A_{i,k} = \binom{n}{k} x_i^k (1 - x_i)^{n-k}$  und  $\mathbf{D}$  die Vorwärtsdifferenz-Matrix ist.

Es wurde sich für eine Bernstein-Polynom-Regression entschieden aufgrund der effizienten Regression und den gering eintretenden Laufzeitkosten bei obfuskierten Programmen. Die Abweichung ...asd

## 6.4 Generierung von ununterscheidbarem Füllcode

Probfuscation übernehmen [1]

## 7 Evaluierung

IDEE: Programmierern das geben und empirisch ermitteln!

### 7.1 Kriterien

Kriterien nach Collberg et al. ...

### 7.2 Vergleich mit existierenden Obfuskationsmethoden

## 8 Ergebnisdiskussion

## 9 Fazit und Ausblick

---

<sup>5</sup>Finden nahe-optimaler Parameter  $c_0, \dots, c_k$ , sodass  $B_n$  die Punkte  $(x_1, y_1), \dots, (x_n, y_n)$  bestmöglich beschreibt

## Literaturverzeichnis

- [1] Juan Caballero, Urko Zurutuza und Ricardo J. Rodríguez, Hrsg. *Detection of Intrusions and Malware, and Vulnerability Assessment: 13th International Conference, DIMVA 2016, San Sebastián, Spain, July 7-8, 2016, Proceedings*. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016. ISBN: 978-3-319-40666-4 978-3-319-40667-1. DOI: 10.1007/978-3-319-40667-1. URL: <https://link.springer.com/10.1007/978-3-319-40667-1> (besucht am 20.07.2025).
- [2] Christian Collberg, Clark Thomborson und Douglas Low. “A Taxonomy of Obfuscating Transformations”. In: <http://www.cs.auckland.ac.nz/staff-cgi-bin/mjd/csTRcgi.pl?serial> (Jan. 1997).
- [3] Christian Collberg, Clark Thomborson und Douglas Low. “Manufacturing Cheap, Resilient, and Stealthy Opaque Constructs”. In: *Proceedings of the 25th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages - POPL ’98*. The 25th ACM SIGPLAN-SIGACT Symposium. San Diego, California, United States: ACM Press, 1998, S. 184–196. DOI: 10.1145/268946.268962. URL: <http://portal.acm.org/citation.cfm?doid=268946.268962> (besucht am 17.07.2025).
- [4] Pascal Junod u. a. “Obfuscator-LLVM – Software Protection for the Masses”. In: *Proceedings of the IEEE/ACM 1st International Workshop on Software Protection, SPRO’15, Firenze, Italy, May 19th, 2015*. Hrsg. von Brecht Wyseur. IEEE, 2015, S. 3–9. DOI: 10.1109/SPRO.2015.10.
- [5] Chris Lattner und Vikram Adve. “LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation”. In: *Proceedings of the International Symposium on Code Generation and Optimization (CGO)*. San Jose, CA, USA, 2004, S. 75–86. DOI: 10.1109/CGO.2004.1281665.
- [6] Hong Lin u. a. “Branch Obfuscation Using Binary Code Side Effects”. In: *Proceedings of the International Conference on Computer, Networks and Communication Engineering (ICCNC 2013)*. The International Conference on Computer, Networks and Communication Engineering (ICCNC 2013). China: Atlantis Press, 2013. DOI: 10.2991/iccnc.2013.37. URL: <http://www.atlantis-press.com/php/paper-details.php?id=6493> (besucht am 13.07.2025).
- [7] “Linear Obfuscation to Combat Symbolic Execution”. In: Zhi Wang u. a. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, S. 210–226. ISBN: 978-3-642-23821-5 978-3-642-23822-2. DOI: 10.1007/978-3-642-23822-2\_12. URL: [https://link.springer.com/10.1007/978-3-642-23822-2\\_12](https://link.springer.com/10.1007/978-3-642-23822-2_12) (besucht am 22.07.2025).
- [8] LLVM Project. *The LLVM Compiler Infrastructure*. <https://llvm.org/>. Version X.Y, accessed 2025-08-18. 2003–2025.
- [9] Ramtine Tofighi-Shirazi u. a. *Defeating Opaque Predicates Statically through Machine Learning and Binary Analysis*. 4. Sep. 2019. DOI: 10.48550/arXiv.1909.01640. arXiv: 1909.01640 [cs]. URL: <http://arxiv.org/abs/1909.01640> (besucht am 23.06.2025). Vorveröffentlichung.
- [10] Hui Xu u. a. “Manufacturing Resilient Bi-Opaque Predicates Against Symbolic Execution”. In: *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Net-

works (DSN). Luxembourg City: IEEE, Juni 2018, S. 666–677. DOI: 10.1109/dsn.2018.00073.  
URL: <https://ieeexplore.ieee.org/document/8416525/> (besucht am 17.07.2025).

## Abbildungsverzeichnis

1	Kontrollflussgraph einer einfachen Funktion mit opakem Prädikat. Abbildung aus der Disassembly des Spiels <i>Overwatch</i> mittels IDA entnommen. . . . .	2
2	Ausschnitt des Kontrollflussgraphen einer Funktion mit vielen opaken Prädikaten. Abbildung aus der Disassembly des Spiels <i>Overwatch</i> mittels IDA entnommen. . . . .	2
3	Konzeptuelles Framework zur Erkennung opaker Prädikate mit symbolischer Ausführung, Abb. aus [10] übernommen . . . . .	2