PoP: Probabilistische opake Prädikate gegen symbolische Ausführung

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

Projektüberblick

Diese Arbeit beschäftigt sich damit, Computerprogramme so zu verändern, dass eine Rekonstruktion ihrer Programmlogik für Angreifer wirtschaftlich unrentabel ist. Hierfür wurde eine Methode entwickelt, die Bedingungen in Programmen einsetzt, um die Komplexität der Programme zu erhöhen, ohne dabei den Programmablauf zu verändern. Im Unterschied zu existierenden Methoden funktioniert diese über die Einbindung von Wahrscheinlichkeiten. Diese erschweren eine Entfernung der eingeführten Bedingungen und somit auch der Komplexität für Angreifer erheblich, wie durchgeführte Experimente belegen.

Inhaltsverzeichnis

1 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, in der Industrie und im Militär[12]. 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. Von besonderem Interesse im Bereich der Obfuskation sind opake Prädikate, eine Kontrollflussobfuskation welche immer wahre bzw. falsche Verzweigungen in Programme einfügt.

In dieser Arbeit wird folgender Frage nachgegangen: *Ist es möglich, opake Prädikate zu kreieren, welche eine automatisierte Deobfuskation mit aktuellen Methoden verhindern und dabei praxisgerecht bleiben?* Die Relevanz dieser Forschungsfrage ergibt sich aus dem fortdauernden Kampf zwischen Obfuskation und Deobfuskation sowie der Effektivität von symbolischer Ausführung gegen opake Prädikate.

Die Kernbeiträge dieser Arbeit hierzu sind folgende:

- Es wird das neue Konzept probabilistischer opaker Prädikate vorgestellt. Es handelt sich dabei um opake Prädikate, welche durch die Nutzung von Wahrscheinlichkeitsverteilungen und Pseudozufallsvariablen verschiedene Angriffsmethoden verhindern.
- Es wird ein Algorithmus geliefert, um mittels Bernsteinpolynomen probabilistische opake Prädikate mit zufälligen Wahrscheinlichkeitsverteilungen zu generieren. Der Algorithmus wird implementiert und erfolgreich evaluiert.
- Dabei wird ein Problem bei der Generierung symbolischer Variablen über Funktionsparameter angesprochen und Algorithmus 4 als Lösung präsentiert. Nach bester Kenntnis des Autors handelt es sich hierbei um einen neuen undokumentierten Angriffsvektor.

In dieser Arbeit wird sich bewusst auf die Erzeugung der opaken Prädikate selbst beschränkt. Die Relevanz von Füllcode (sog. *Junkcode*) ist unabhängig von der dargestellten Idee und wird daher nicht behandelt. Zunächst werden Obfuskation, opake Prädikate und symbolische Ausführung formal definiert (Abschnitt 2). Durch einen Überblick über den aktuellen Stand der Forschung (Abschnitt 3) wird die Notwendigkeit dieser Arbeit hergeleitet und ein Angreifermodell festgelegt (Abschnitt 4.1). In Abschnitt 4.2-3 wird die Idee propabilistischer opaker Prädikate vorgestellt und definiert. Darauf werden verschiedene Methoden zur zufälligen Wahrscheinlichkeitsverteilungsgenerierung (Abschnitt 4.4) sowie zur Generierung von Pseudozufallszahlen (Abschnitt 4.5) vorgestellt und abgewogen. Dabei wird das bis jetzt nicht untersuchte Problem der Aufruf-Invarianz für die Generierung symbolischer Variablen über Funktionsparameter angesprochen und über Algorithmus 4 gelöst. Implementiert wird die Idee in Form eines LLVM-Passes (Abschnitt 5). Eine Evaluierung anhand der Kriterien Collbergs et al. [8] (Abschnitt 6) zeigt, dass probabilistische opake Prädikate resilient gegenüber symbolischer Ausführung und weiteren Methoden sind, ohne dabei unvertretbare Kosten aufzuweisen.

Diese Arbeit geht von einem informatischen Grundwissen an Assembler, Compilern sowie Programmanalysemethoden aus. Zudem wird die Iverson-Klammer/ Prädikatabbildung $[\cdot]$ verwendet: Unter der Voraussetung, dass die Aussage P wahr ist, gilt $[P] = 1$. Ansonsten gilt $[P] = 0$. Pseudocode, welcher Programmanweisungen modifiziert nimmt an, dass diese in *Single Static Asignment*-Form definiert sind.

2 Theoretische Grundlagen

2.1 Obfuskation

Collberg et al. [8] definieren Obfuskation wie folgt:

Definition 2.1 (Obfuskation)**.** Sei $P \xrightarrow{\mathcal{T}} P'$ eine Transformation $\mathcal{T}$ eines *Quellprogrammes* P zu einem *Zielprogramm* P'. Eine solche Transformation ist eine Obfuskation, wenn das obfuskierte Programm P' dasselbe beobachtbare Verhalten wie P für den Endnutzer aufweist.

Obfuskation zielt darauf ab, die Komplexität eines Programmes so zu erhöhen, dass dessen interne Logik für einen Angreifer nur schwer verständlich ist. Per Definition sind 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 aus, wie später ersichtlich sein wird.

Das Rückgängigmachen einer Obfuskation ist die *Deobfuskation*.

2.2 Opake Prädikate

Die folgenden Definitionen sind aus [27] sowie vom Pionierwerk [9] modifiziert übernommen. Es wird sich auf hierbei auf invariante opake Prädikate beschränkt.

Definition 2.2 (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 sein 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 [9].

Abbildung 1: Kontrollflussgraph einer einfachen Funktion mit opakem Prädikat.

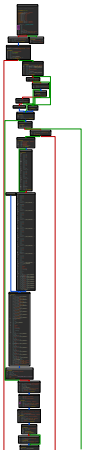

Abbildung 2: Ausschnitt des Kontrollflussgraphen einer Funktion mit vielen opaken Prädikaten. Durch die vielen opaken Prädikate wird die Funktion unübersichtlich und eine Analyse aufgrund der Unklarheit tatsächlich ausführbarer Pfade erschwert.

Opake Prädikate werden in der Softwareobfuskation eingesetzt, um ein Verständnis über den Kontrollfluss des Programms zu behindern [9, 27]. Damit opake Prädikate als Obfuskationsmethode [1] 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 [9]. Mit zunehmender Komplexität der Prädikate und zunehmender Anzahl dieser, nimmt also auch die Obfuskationsstärke (Verwirrung und Unverständnis) beim Angreifer zu (vgl. Abb. 2).

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* $O(\phi) = [(y < 10) \vee (x \cdot (x + 1) \mod 2 \equiv 0)]$ *aus [14] mit* $\phi = (x, y)$ *und* $x, y \in \mathbb{Z}$ *ist immer wahr, da* $x \cdot (x + 1)$ *immer gerade ist. Der Wert is folglich von y unabhängig.*

[1] D.h, dass der wirkiche Pfad, welcher von einem opaken Prädikat verschleiert wird, nicht einfach erkannt werden kann

2.3 Symbolische Ausführung[2]

Im Gegensatz zur konkreten Ausführung, welche ein Programm für spezifische Inputs ausführt, ermöglicht die symbolische Ausführung die Analyse des Programmverhaltens für ganze Klassen an Inputs. Die Notwendigkeit symbolischer Ausführung ergibt sich schon am Beispiel einer einer Funktion mit zwei 64-Bit Variablen. Um mit dynamischer (konkreter) Ausführung herauszufinden, für welche Werte eine Bedingung wahr ist, müsste man hier $2^{64} \cdot 2^{64} = 2^{128}$ verschiedene Werte ausprobieren. Ein solcher Bruteforce ist selbst für die modernsten Computer unmöglich - symbolische Ausführung hingegen schon. Ein symbolischer *Ausführungsengine* besteht aus 2 Hauptkomponenten: einem *symbolischen Ausführungsmodul* und einem *Constraint-Solver*[3] zur Lösung/zum Prüfen von Bedingungen/Einschränkungen. Bei der symbolischen Ausführung wird für jeden Kontrollflussweg eine *Pfadformel* und ein *symbolischer Speicher* mitgeführt.

1. Die Pfadformel, eine boolesche Formel erster Ordnung, führt die Bedingungen der entlang des Pfades genommenen Verzweigungen zusammen.

2. Der symbolische Speicher bildet unbekannte Variablen (z.B. Parameter und alle darauf aufbauende Variablen) auf symbolische Ausdrücke ab.

Hierdurch können schließlich über den *Constraint-Solver* allgemeine Aussagen über die Erreichbarkeit bestimmter Pfade oder Variablenwerte getroffen werden. Ist eine Pfadformel erfüllbar, kann der Solver zudem konkrete Eingabewerte hierfür liefern. Hat das Programm aber besonders viele Verzweigungen (z.B. durch Schleifen) oder komplexe Constraints (z.B. nichtlineare Arithmetik), stoßen die *Constraint-Solver* an ihre laufzeittechnischen Grenzen. Zur Lösung wurden verschiedene Ansätze (z.B. *Concolic Execution*) entwickelt. Aufgrund der Fülle an Informationen und der geringen Relevanz für die in dieser Arbeit dargestellten Abwehrmethodik, wird auf ihre Darstellung verzichtet.

3 Stand der Forschung und Motivation

Dieser Abschnitt präsentiert den aktuellen Stand der Forschung zu opaken Prädikaten und begründet daraus diese Arbeit. Es werden aktuelle, zentrale Ansätze exemplarisch vorgestellt, um Forschungsstand und Herausforderungen zu verdeutlichen. Die geschieht anhand der Kriterien aus Abschnitt 6.1.

Existierende Literatur beschränkt sich vornehmlich auf statische Analyseansätze. Dynamische Analyseideen z.B. zur probabilistischen Untersuchung opaker Prädikate wurden veröffentlicht und experimentell untersucht, ergaben aber eine zu hohe Fehlerquote. Insbesondere reduzieren sich publizierte Ansätze auf symbolische Ausführung. Dies hat den Hintergrund, dass die symbolische Ausführung zur Zeit eine der effektivsten automatisierten Analysemethoden ist, welche mit wenig Aufwand/manuellen Eingriffen verwendet werden kann. Andere Analysemethoden, wie z.B. *Tainting* sind zudem abhängiger von Faktoren neben dem opaken Prädikat selbst. Im Falle des *Taintings* ist die Qualität des Füllcodes wesentlich [31].

Existierende Methoden lassen sich im Wesentlichen in zwei Kategorien einteilen:

1. Opake Prädikate, welche Implementierungsschwächen[4] existierender symbolischer Ausführungsengines angreifen.

   Hierunter fällt die Nutzung von *Exceptions* [16], Anweisungen [30] oder Funktionen [7], welche nicht durch die symbolischen Ausführungsengines modelliert werden. Eine Befragung von Audrey Dutcher, einer der Entwicklerinnen vom Programmanalyseframework *Angr* [25] ergab: drei von vier der in [30] dargestellten Methoden können nun von dem symbolischen Ausführungsengine von *Angr* problemlos symbolisch

[2] Die Angaben des folgenden Abschnitts entstammen, sofern nicht anders vermerkt, aus [1]
[3] Es handelt sich meist um einen SMT-Solver.
[4] bzw. Heuristikschwächen.

ausgeführt werden [5].Eine Deobfuskation weiterer Methoden in dieser Kategorie liegt also alleine in der Verbesserung existierender symbolischer Ausführungsengines. Eine Deobfuskation ist in gewisser Frage nur eine Frage der Zeit.

2. Opake Prädikate, welche fundamentale Grenzen der *Constraint Solver* angreifen.

   Diese nutzen entweder ungelöste Probleme in der Mathematik oder NP-schwere Probleme der Informatik. So erstellte [29] z.B. opake Prädikate, welche auf der unbewiesenen Vermutung basieren, dass die Collatz Folge unabhängig vom Startwert immer den Zyklus $4, 2, 1$ erreicht. Da Mathematiker noch nicht die Werkzeuge haben, dies zu beweisen, ist die symbolische Ausführung auch nicht in der Lage zu beweisen, dass das Programm weiterläuft und nicht in eine Endlosschleife gerät. In [21] wird alternativ ein NP-schweres Problem mit Pointerarithmetik konstruiert. In beiden Fällen sind die opaken Prädikate zwar beweisbar sicher, allerdings einfach erkennbar und mit hohen Laufzeitkosten verbunden. Andere Methoden wie opake Prädikate mit *Mixed Boolean Arithmetic* Ausdrücken lassen sich mit neusten Heuristiken mitlerweile trotz NP-schwere größtenteils deobfuskieren [23].

Angesichts der dargelegten Probleme aktueller Methoden ergibt sich die Notwendigkeit effizienter, getarnter und resilienter (nur schwer automatisch deobfuskierbarer) opaker Prädikate.

4 Ansatz

4.1 Angreifermodell

Diese Arbeit geht aufgrund der Ähnlichkeit behandelter Thematik von einem *Man-at-the-End* (MATE) Angreifermodell aus, aufbauend auf [30, 31]. Ein Angreifer hat direkten Zugriff auf das Programm und dessen Anweisungen sowie volle Kontrolle über das Endsystem, auf dem sie ausgeführt werden.
Es ist dem Angreifer hierbei nicht vorgegeben, wo und inwiefern das Programm obfuskiert ist. Der Angreifer kann das Programm statisch und dynamisch analysieren. Das Ziel des Angreifers ist dabei, ein Verständnis der obfuskierten Programmlogik zu gewinnen. Pattern Matching, also das Suchen von Assembler-Anweisungsfolgen, kann hierbei zum Finden und Löschen zuvor erkannter opaker Prädikate verwendet werden. Zudem kann der Angreifer Funktionen symbolisch ausführen. Über eine Anfrage an den Constraint-Solver kann hierbei geprüft werden, ob ein Prädikat für alle Eingabewerte wahr ist. Ist dies der Fall, so handelt es sich um ein opakes Prädikat, welches gelöscht werden kann. Mögliche dynamischer Analysemethoden werden in Abschnitt 7 angeführt und diskutiert.

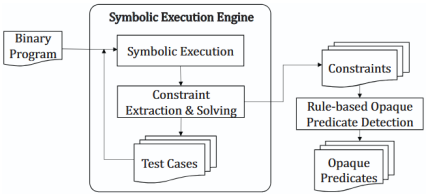

Abbildung 3: Konzeptionelles Framework zur Erkennung opaker Prädikate mit symbolischer Ausführung. Abbildung aus [30] übernommen.

4.2 Probabilistische opake Prädikate

Symbolische Ausführung genügt für die Untersuchung deterministischer Algorithmen und entscheidbarer Probleme. Will man aber Aussagen über einen probabilistischen Algorithmus treffen, so ist dies ohne erhebliche manuelle Eingriffe eines Nutzers in Form von extra Annahmen und Beschränkungen (*Constraints*) unmöglich. Jede zufällige Variable eines jeden Schleifenaufrufes muss vom Constraint Solver als symbolische Variable betrachtet und das Programm für alle möglichen Werte untersucht werden. Bei einem Monte-Carlo-Algorithmus bedeutet dies, dass auch

[5](a) symbolischer RAM: Ausführbar für Arrays mit einer Länge unter 257.
(b) Gleitkommazahlen: Ausführbar, wenn keine x86 `long double` Datentypen verwendet werden.
(c) Verdeckte symbolische Kontrollflussübertragung (*Covert Symbolic Propagation*), in [30] über Dateisystem-Operationen implementiert: Ausführbar.
(d) Threads: Noch nicht implementiert.

unwahrscheinliche Variablenwerte, welche zu falschen Ergebnissen führen, überprüft werden. Der Wahrheitsgehalt wird somit zwar formal logisch korrekt bewiesen - praktisch allerdings nicht.

Beispiel 3. *Man betrachte einen Algorithmus, welcher mit einer Wahrscheinlichkeit von* $99,9999\%$ *den Wert* 1 *zurückgibt und mit einer Wahrscheinlichkeit von* $0,0001$ *den Wert* 0 *zurückgibt.*

Algorithm 1 Beispiel eines probabilistischen Algorithmus

```
1: procedure Foo()
2:     X ← UniformRand(0, 1)
3:     if X ≤ 0,999999 then
4:         return 1
5:     end if
6:     return 0
7: end procedure
```

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

Dies bildet die Grundidee probabilistischer 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 4.1 (Probabilistische opake Prädikate)**.** Sei $O_p : \Phi \to \{0,1\}$ eine Abbildung einer Variable $\phi \in \Phi$ zu einem Prädikat. Das Prädikat $O_p(\phi)$ ist probabilistische opak, wenn der Fall $A \in \{0,1\}$ mit einer so hohen Wahrscheinlichkeit eintritt, dass der Fall $\overline{A}$ vernachlässigbar ist.

Definition 4.2. Sei X eine Zufallsvariable mit beliebiger Wahrscheinlichkeitsverteilung $P(X;\theta)$, wobei θ die Parameter der Verteilung darstellt. Das probabilistische opake Prädikat $O_p(\phi)$ wird wie folgt konstruiert:

$$O_p(\phi) = [f(X) \bowtie c]\,, \tag{1}$$

wobei:

1. f eine Transformation durch arithmetische und bitweise Operationen ist (z. B. $f(X) = m \cdot (X \oplus k) + b)$), mit Konstanten $(m, k, b) \in \mathbb{R}$,
2. $\bowtie$ ein Zahlenvergleichsoperator ist $(=, >, <, \geq, \leq)$ und
3. c eine festgelegte Konstante ist.

Definition 4.3. $\mathcal{P}_{prob}$ ist die Klasse aller probabilistische opaker Prädikate.

Beispiel 4. *Ein einfaches probabilistisches opakes Prädikat ist* $O_p(\phi) = \left[\textsc{Uniform}(0,1) \leq 1 - 10^{-2}\right]$. *Die Wahrscheinlichkeit, dass dieses Prädikat wahr ist, beträgt* $1 - 10^{-2} = 0,99\%$.

Beispiel 5. $O_p(\phi) = [\textit{Poisson}(5) \geq 15]$. *Die Wahrscheinlichkeit, dass dieses Prädikat unwahr ist beträgt* $\sum_{k=15}^{\infty} \frac{5^k e^{-5}}{k!} = 1 - \sum_{k=0}^{14} \frac{5^k e^{-5}}{k!} \approx 0,023\%$

Um zu garantieren, dass das Programm, in welchem das probabilistische opake Prädikat eingefügt wurde, weiterhin funktioniert, kann für den ungewünschten Gegenfall eine Wahrscheinlichkeit genutzt werden, welche unter der eines Hardwarefehlers liegt.

Die geringe Wahrscheinlichkeit, dass die Prädikate in $\mathcal{P}_{prob}$ sich nicht wie gewünscht verhalten, gewährleistet ihnen theoretische Resistenz gegenüber symbolischer Ausführung. Ohne Heuristiken ist es (bei adäquater Implementierung) theoretisch unmöglich, zwischen einem „normalen“ Prädikat, welches besonders häufig einen Wert annimmt, und einem probabilistischen opaken Prädikat zu unterscheiden. (vgl. Beispiel 6) Dies zwingt den Angreifer zu einer genaueren Analyse jedes Prädikats im Sachzusammenhang.

Beispiel 6. *Sei p ein Prädikat, welches zu* 95% *der Zeit wahr ist. Ist* $p \in \mathcal{P}_{prob}$ *oder einfach besonders häufig wahr?*

4.3 Konstruktionsalgorithmus

Algorithm 2 Generierung probabilistischer opaker Prädikate

Require: D_{start}, D_{end} (Domain of CDF), $P(RealBB)$ (Probability of generated predicate having wanted value), *Precision* (determines *Threshold* precision in predicate)

1: **procedure** GeneratePROBABILISTICPREDICATES($Function\ F \in Module, AlwaysTruePred \in \{0,1\}, Precision \in \mathbb{N}$)
2: $BB \leftarrow$ GetRandomBasicBlock(F)
3: $U \leftarrow$ GetSymbolicVariable().AsInt64() ▷ Generate a random variable.
4: $U \leftarrow$ BB.InsertMultiplication(U, GetRandomOddNr()) ▷ Hashing: $U \sim Unif$.
5: $U \leftarrow$ BB.InsertDivision(U, INT_MAX) ▷ $U \in [0;1]$.
6: $CDF \leftarrow$ GenerateRandomDistribution()
7: BB.InsertCallInverseCDF(CDF, U)
8: $Threshold \leftarrow \frac{D_{end}-D_{start}}{2}$
9: **for** $i \leftarrow 0$ **to** *Precision* **do** ▷ Compute threshold using the Newton–Raphson method.
10: $OffsetY \leftarrow$ CDF.EvaluateAt($Threshold$) $- P(RealBB)$
11: $Slope \leftarrow$ CDF.EvaluateDerivativeAt($Threshold$)
12: $Threshold \leftarrow Threshold - \frac{OffsetY}{Slope}$
13: $Threshold =$ max(min($Threshold, D_{start}$), D_{end}) ▷ Clamp to prevent divergence.
14: **end for**
15: $RealBB \leftarrow$ BB.Split($LastInstruction$) ▷ Always executed Basicblock
16: $FakeBB \leftarrow$ F.CreateBB() ▷ Never executed Basicblock
17: FakeBB.InsertJunkCode()
18: **if** *AlwaysTruePred* **then**
19: BB.InsertIf($U < Threshold, RealBB, FakeBB$)
20: **else**
21: BB.InsertIf($U > Threshold, FakeBB, RealBB$)
22: **end if**
23: **end procedure**

Für jedes zu generierende opake Prädikat wird im Programm eine Pseudozufallsvariable U generiert. Verschiedene Methoden hierfür werden in Abschnitt 4.5 gegeben. Wichtig ist, dass der Angreifer den Wert von U nicht statisch bestimmen kann. Um die Gleichverteilung von U zu erhalten und die Bits zu verstreuen wird die Zufallsvariable mit einer zufälligen ungeraden Konstante modulo 2^{64} multipliziert. Da jede ungerade Zahl teilerfremd zu 2^{64} ist, ist diese Abbildung bijektiv.

Aus den probabilistischen opaken Prädikaten ohne Transformationen ($f(X) = X$) lassen sich Wahrscheinlichkeitswerte direkt herauslesen (vgl. Beispiel 4). Um dies zu vermeiden wird über die sog. Inversionsmethode U in eine andere z.B. normal-, exponential- oder bernoulliverteilte Zufallsvariable transformiert. Wichtig ist dabei, dass diese neue Verteilung in „wahrscheinliche“ bzw. „unwahrscheinliche“ Intervalle eingeteilt werden kann. Das Vorgehen hierfür wird in Abschnitt 4.4 genauer vorgestellt. Mit dieser transformierten Zufallsvariable können nun wahrscheinliche bzw. unwahrscheinliche probabilistische Prädikate erstellt werden.

Der Pseudocode für die Generierung solcher probabilistischer opaker Prädikate ist in Algorithmus 2 beschrieben.

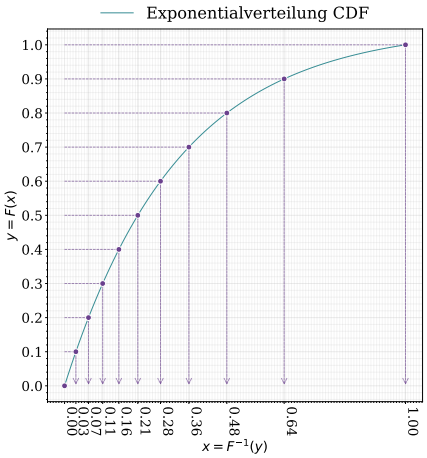

Abbildung 4: Zufällige Zahlen y_i werden von einer Gleichverteilung $Unif(0,1)$ generiert. Jeder zufällige y-Wert wird über die inverse kumulative Verteilungsfunktion der Exponentialverteilung $F^{-1}(x)$, einem x-Wert zugeordnet. Wie bei einer Exponentialverteilung sammeln sich hierdurch die $x = F^{-1}(x)$-Werte um 0.

4.4 Wahrscheinlichkeitsverteilungsgenerierung

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

Generierung über Verteilungsfamilie Eine Möglichkeit ist die Nutzung einer Wahrscheinlichkeitsverteilung, deren Wahrscheinlichkeits-(dichte-)funktion über einen oder mehrere Parameter bestimmt wird.

Ein Beispiel hierfür ist die Gammaverteilung mit Skalenparameter $\alpha > 0$, Formparameter $r > 0$ und folgender Dichtefunktion [13][6]:

$$\gamma_{\alpha,r}(x) \;=\; \frac{\alpha^r}{\Gamma(r)} x^{r-1} e^{-\alpha x}, x > 0. \tag{2}$$

Diese hat den Vorteil, dass sich verschiedene andere Verteilungen (z.B. Chi-Quadrat-, Erlang und Exponentialverteilung) aus ihr ergeben [13]. Ein Pattern Matching Angriff müsste demnach nach Termen der sehr allgemeinen Form $ax^n b^x$ suchen[7]. Dennoch ist es nicht unmöglich: Es ist nicht davon auszugehen, dass der Term in regulären Programmen (oft/überhaupt) vorkommt.

Generierung über zufälliges Polynom Stattdessen wurde sich entschieden für die Generierung einer zufälligen Funktion F, welche die Eigenschaften einer kumulativen Verteilungsfunktion erfüllt. Eine kumulative Verteilungsfunktion $F : \mathbb{D} \to [0; 1]$ muss folgende 3 Eigenschaften erfüllen[13]:

1. F ist monoton steigend.
2. F ist rechtsseitig stetig.
3. $\lim\limits_{x \to \inf \mathbb{D}+} F(x) = 0$ und $\lim\limits_{x \to \sup \mathbb{D}-} F(x) = 1$.

Die Umsetzung über Polynome in der Standardbasis ist durch deren häufigen Oszillationen erschwert. Der einfachste Weg, Monotonie sowie die beschriebenen Grenzwerte umzusetzen, ist über Bernsteinpolynome folgender Form[19]:

$$B_n(x) \;=\; \sum_{k=0}^{n} c_k \binom{n}{k} x^k (1-x)^{n-k}, \text{mit } c_0 \le c_1 \le \cdots \le c_n \text{ und } n \in \mathbb{R}. \tag{3}$$

Um alle Eigenschaften einer kumulativen Verteilungsfunktion zu erfüllen, muss $B_n(x)$ zudem durch die Punkte $(0|0)$ und $(0|1)$ verlaufen. Hierfür genügt es, $c_0 = 0$ und $c_n = 1$ festzulegen [19].

Für eine Verallgemeinerung lassen sich der horizontaler Skalierungsparameter a sowie der horizontale Verschiebungsparameter k in $B_n(a \cdot (x - k))$ einfügen.

Das Vorgehen zur Generierung einer CDF mit Bernsteinpolynomgrad n ist somit Folgendes:

1. Wähle zwei zufällige rationale Zahlen $(a, k) \in \mathbb{R}$.
2. Definiere den Definitionsbereich $\mathbb{D} = [x_1, x_2]$ mit $x_1 = k$ und $x_2 = k + \frac{1}{a}$.
3. Teile $[x_1; x_2]$ in n Teile ein.
4. Ziehe $n - 1$ unabhängige, gleichverteilte Werte $(c_1, \ldots, c_n) \sim U(0, 1)$ und sortiere sie:

$$c_1 \le c_2 \le \cdots \le c_{n-1}$$

5. Setze $c_0 = 0$ und $c_n = 1$.
6. Generiere die Funktion $B_n(x) = \sum_{k=0}^{n} c_k \binom{n}{k} x^k (1-x)^{n-k}$ mit gegebenen Koeffizienten $(c_0 \ldots c_n) \in [0; 1]$.

[6] wobei $\Gamma(n) = (n-1)!$ ist.
[7] Die Faktoren werden im Vorhinein zur Kompilationszeit miteinander multipliziert

Das Inverse der kumulativen Verteilungsfunktion lässt sich effizient über das Newton-Raphson-Verfahren mit Laufzeitkomplexität $O(k)$ berechnen, wobei k die Anzahl an Iterationen festlegt. Hierfür muss das Bernsteinpolynom mit *streng* monoton steigenden Koeffizienten (Mindestabstand) generiert werden, da die mangelende Steigung sonst zu einer Division durch 0 führen würde (vgl. Algorithmus 3). Aufgrund des quadratischen Konvergenzverhaltens des Verfahrens genügen für 64-Bit Gleitkommazahlen praktisch $k = 6$ Iterationen: Gemäß der IEEE 754 Norm haben 64-Bit Gleitkommazahlen eine 52-Bit Mantisse. Durch das Konvergenzverhalten verdoppelt sich die Anzahl korrekter Bits mit jeder Iteration. Ist anfänglich 1 Bit korrekt erraten, so braucht es 6 Iterationen, um alle $2^6 = 64 > 53$ Bits der Mantisse korrekt zu bestimmen. Um Sicherheitsgarantien zu liefern wurde der Algorithmus mit $k = 7$ implementiert, um bei schlechten anfänglichen Schätzungen und ungünstigen Bernsteinpolynomen dennoch garantiert das Inverse der kumulativen Verteilungsfunktion korrekt zu bestimmen. Die zusätzlichen Laufzeitkosten hierdurch sind minimal. Um eine mögliche Divergenz des Newton-Raphson-Verfahrens zu verhindern, wird der geschätzte Wert bei jeder Iteration auf den Definitionsbereich $\mathbb{D}$ der Funktion begrenzt. Der Pseudocode hierfür ist in Algorithmus 3 gegeben.

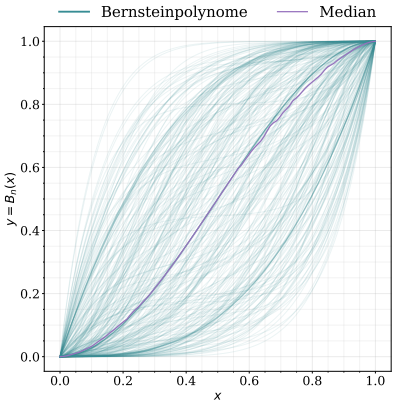

Abbildung 5: Zufällige Bernsteinpolynome fungieren als verschiedene kumulative Verteilungsfunktionen für die probabilistischen opaken Prädikate.

Algorithm 3 Berechnung inverser kumulativer Verteilungsfunktion über das Newton-Raphson-Verfahren

Require: C (Bernstein coefficients), *HShift*, *VStretch* (horizontal shift/ vertical stretch), *Precision* (determines precision of Bernstein evaluation)

function SAMPLEBERNSTEININVERSE($U \in [0, 1]$)
  $Estimate \leftarrow HShift + \frac{0.5}{VStretch}$ ▷ Start guess at center of domain.
  $D_{start} \leftarrow HShift$
  $D_{end} \leftarrow HShift + \frac{1}{VStretch}$
  for $i \leftarrow 0$ **to** *Precision* **do**
    $t \leftarrow VStretch \cdot (Estimate - HShift)$ ▷ Transformed *Estimate*.
    $y \leftarrow 0.0$ ▷ Accumulator for $B(t)$.
    $y' \leftarrow 0.0$ ▷ Accumulator for $B'(t)$.
    for $k \leftarrow 0$ **to** *Degree* **do** ▷ Evaluate CDF $B(t)$
      $b_k \leftarrow \binom{n}{k} \cdot t^k \cdot (1-t)^{n-k}$
      $y \leftarrow y + C[k] \cdot b_k$
    end for
    for $k \leftarrow 0$ **to** *Degree* $- 1$ **do** ▷ Evaluate PDF.
      $b'_k \leftarrow \binom{n-1}{k} \cdot t^k \cdot (1-t)^{n-1-k}$
      $y' \leftarrow y' + C'[k] \cdot b'_k$
    end for
    $OffsetY \leftarrow y - U$
    $Slope \leftarrow y' \cdot HStretch$ ▷ Apply chain rule: $\frac{dy}{dx} = \frac{dy}{dt} \cdot \frac{dt}{dx}$.
    $Estimate \leftarrow Estimate - \frac{OffsetY}{Slope}$ ▷ Newton Step.
    $Estimate = \text{MAX}(\text{MIN}(Estimate, D_{start}), D_{end})$ ▷ Clamp to prevent divergence.
  end for
  return *Estimate*
end function

4.5 Generierung von Pseudozufallsvariablen

Damit der vorgestellte Ansatz funktionieren kann, bedarf er einer gleichverteilten (Pseudo-)Zufallszahl, welche auf dem Einheitsintervall $[0; 1]$ liegt und zugleich als unbestimmte symbolische Variable ohne konkreten Wert von symbolischen Ausführungsengines betrachtet wird. Dafür wird für jede zu obfuskierende Funktion ein zufälliger

Funktionsparameter ausgewählt und auf dem Einheitsintervall abgebildet. Dadurch sind die opaken Prädikate gut getarnt – ein Parameterzugriff ist schließlich normales Verhalten in jedem Programm. Die einzige Gefahr ist dabei, dass der gewählte Parameter immer konstante Werte annimmt. Ein Angreifer könnte in diesem Fall alle Werte dieses Parameters in Funktionsausrufen sammeln und die Funktion mit ihnen ausführen. Sind für jeden Parameterwert die ausgeführten Basisblöcke gleich, so handelt es sich bei den nicht ausgeführten Basisblöcken um „Junkcode“. Prädikate, welche auf diese Basisblöcke verweisen sind folglich opak und können gelöscht werden. **Diese Vulnerabilität wurde in keiner der gelesenen Arbeiten zum Thema bis jetzt angesprochen.** Algorithmus 4 bietet hierfür eine Lösung.

Algorithm 4 Suche nach Parameter mit extern-abgeleiteten Wert in Funktionsaufruf über DFS

```
1:  function GetSymbolicVariable(Function F ∈ Module)
2:      for all Callsite CS ∈ F do
3:          for all Argument Arg ∈ CS do
4:              Visited = HashMap()
5:              if IsVariableDerivedFromExternal(Arg, Visited) then return Arg
6:              end if
7:          end for
8:      end for
9:  end function
10:
11: function IsDerivedFromExternal(Argument Arg, Hashmap Visited, Depth ∈ ℤ)
12:     if Depth > 30 ∨ Arg ∈ Visited then return false        ▷ Constant was deemed empirically suitable.
13:     end if
14:     Visited[Arg] = true
15:     if Val is result of External Function then return true        ▷ Check immediate origin
16:     end if
17:     if Val is Internal Function Call then        ▷ Determine data sources S based on operation type
18:         S ← Return statements of the called function
19:     else if Val is Memory Read then
20:         S ← Values written to the source address (by Stores)
21:     else        ▷ Arithmetic, Casts, PHI inputs
22:         S ← Val.Operands
23:     end if
24:     for all Src ∈ S do        ▷ Recursively trace data flow
25:         if IsDerivedFromExternal(Src, Visited, Depth + 1) then return true
26:         end if
27:     end for
28:     return false
29: end function
```

5 Implementierung

Zur Implementierung wurden drei Ansätze erwogen: die Entwicklung eines Bin2Bin-Obfuskators[8], die Implementierung in Form eines LLVM-Passes [15, 17][9] sowie die Quellcodemanipulation[10] Es wurde eine Entscheidung für einen LLVM-Pass getroffen aufgrund folgender Vorteile:

1. **Abstraktion und Portabilität:** LLVM entkoppelt durch eine abstrakte Zwischensprache (*Intermediate Representation*) von Architektur-/Betriebssystemdetails. Viele *low-level* Aufgaben (z.B. *Relocations*, Einfügen von Assembler-Anweisungen etc.) werden übernommen.

2. **Optimierung:** Die LLVM-Toolchain enthält etablierte Optimierungs-Pässe und profitiert fortlaufend von der Arbeit zahlreicher Beitragender. Dies ermöglicht eine nahezu optimale Kompilation obfuskierter Programme.

[8] Direkte Manipulation von Assembler-Anweisungen existierender Programme.
[9] Nutzung des LLVM-Projekts, um einen sog. *Compiler-Pass* zu schreiben.
[10] durch z.B. C-Makros und das Einfügen von inline Assembler-Ausschnitten

Die ist nützlich/erforderlich in vielen Anwendungsszenarien (z.B. *Embedded Systems*, IoT, Echtzeitsysteme etc.).

3. **Entwicklungsaufwand:** Bin2Bin-Obfuskatoren und umfangreiche Quellcodemanipulation erfordern viel manuellen Aufwand und sind fehlerhaftig. Ein LLVM-Pass ermöglicht den reinen Fokus auf die Obfuskationslogik.

Ein LLVM-Pass bietet in diesem Fall das beste Kompromissverhältnis zwischen *low-level* Kontrolle, Laufzeiteffizienz, einer *high-level* Portabilität sowie vertretbarem Entwicklungsaufwand.

Der Pass wurde mit mehreren Parametern versehen (vgl. Abb. 6a): Nutzer können angeben, wie viel Prozent der Funktionen obfuskiert werden sollen (`-pop-lvl`), wie wahrscheinlich der korrekte Pfad der probabilistischen opaken Prädikate sein soll (`-pop-pred-prob`) sowie den minimalen/maximalen Bernsteinpolynomgrad (`-pop-min/max-degree`). Zudem können Funktionen im Quellcode als zu obfuskierend markiert werden (vgl. Abb. 6b)

Die Implementierung, Beispiele und Experimente sind unter [4] veröffentlicht.

```bash
#!/bin/bash

OPT_LVL=O3
FILENAME="hello_world"
PASS_PLUGIN_DIR="Obfuscator.so"
LVL=50 RUNS=1 PRED_PROB=99
MIN_DEG=3 MAX_DEG=6

clang -$OPT_LVL \
-fpass-plugin=$PASS_PLUGIN_DIR \
-Xclang -load -Xclang $PASS_PLUGIN_DIR \
-mllvm -pop-lvl=$LVL \
-mllvm -pop-pred-prob=$PRED_PROB \
-mllvm -pop-runs-per-fn=$RUNS \
-mllvm -pop-min-degree=$MIN_DEG \
-mllvm -pop-max-degree=$MAX_DEG \
${FILENAME}.c \
-o $FILENAME
```

(a) Bash-Skript zur Ausführung der implementierten Obfuskation auf C(++) Quellcode.

```c
__attribute__(annotate("POP"))
void foo(int x)
{
  printf("foo %i\n", x);
}
```

(b) Eine einfache zu obfuskierende Funktion.

```c
void __fastcall foo(uint64_t x)
{
  __m128d v7;
  uint64_t v8 = x;
  double v9 = x * COERCE_DOUBLE(0x4000000000000LL);
  double v10 = 35.31442506435751;
  for ( i = 0; i != 7; ++i )
  {
    __m128d v1 = *&v10;
    double v6 = (v10 - 35.299408635128881) * 33.29686388054899;
    v1.m128d_f64[0] = v10 - (0.0 * ((1.0 - v6) * (1.0 - v6) *
     (1.0 - v6)) + 0.5018965731323943 * v6 * ((1.0 - v6) * (1.0 -
      v6)) + 1.199789464169086 * (v6 * v6) * (1.0 - v6) + v6 * v6
      * v6 - v9) / ((0.5018965731323943 * ((1.0 - v6) * (1.0 - v6
     )) + 1.395785782073383 * v6 * (1.0 - v6) + 1.800210535830914
      * (v6 * v6)) * 33.296863880548899);
    v7 = v1;
    v2 = _mm_cmplt_sd(0x4041AA2B238C72F8uLL, v1);
    v3 = _mm_or_pd(_mm_andn_pd(v2, v1), _mm_and_pd(v2, 0
     x4041AA2B238C72F8uLL));
    v4 = _mm_cmplt_sd(v3, 0x4041A65305AC0256uLL).m128d_f64[0];
    *&v10 = ~*&v4 & *&v3.m128d_f64[0] | *&v4 & 0
     x4041A65305AC0256LL;
  }
  if ( v7.m128d_f64[0] <= 35.32927409341327 )
    printf("foo %lu\n", *(&v5 - 2)); // This Basic Block will
     always be executed.
  else
    // Junkcode; This Basic Block will never be executed
}
```

(c) Dieselbe Funktion aber mit PoP (Bernsteinpolynomgrad $n = 3$) obfuskiert. Pseudo-C wurde modifiziert aus der Dekompilation von IDA entnommen. Die Funktionen mit dem Unterstrich als Präfix sind intrinsische Funktionen. Sie kapseln einzelne CPU-Anweiseungen.

6 Evaluierung

6.1 Vorgehen

Für die Evaluierung wurden die weitverbreiteten Kriterien von Collberg et al. [8] verwendet: Es wurde sich bewusst

Kriterium	Beschreibung
Kosten	Wie sehr erhöht die Obfuskation die Laufzeitkosten?
Stärke	Wie unverständlich ist das obfuskierte Programm für einen Angreifer?
Resilienz	Wie schwer ist eine (automatisierte) Deobfuskation?
Tarnung	Wie auffällig ist die Obfuskation?

Tabelle 1: Kriterien der Obfuskation

gegen die einfachen Benchmarkprogramme wie in [2, 3] oder [22] entschieden. Orientiert wurde sich dabei am

zur Zeit umfangreichsten Literaturreview zur (De-)Obfuskation [26]. Dieses bemängelt mangelnde Programmgröße sowie -quantität in der aktuellen Obfuskationsforschung. Es wurde sich daher für die LLVM Test Suite [18] entschieden. Dabei wurde sich auf die größten vollständigen Anwendungen (darunter z.B. SQLite & Lua) und Benchmarkprogramme der `MultiSource` Untereinheit (31 Programme) beschränkt. Obwohl die Test Suite eigentliche für Compilerevaluierungen entwickelt wurde, ist sie dennoch für die Obfuskationsevaluierung geeignet, da sie durch ihre diversen Programme realistische Anwendungsszenarien der Obfuskation abbildet. Die Test Suite bietet zudem den Vorteil vorgefertigter Tests (167 Tests), welche die Korrektheit obfuskierter Programme prüfen.

Alle Programme wurden mit clang 18.1.3 und Optimierungslevel `-O3` kompiliert und mit Ubuntu 24.04.3 LTS mit Kernel Version 5.16, einer Intel® Core™ i7-10700F CPU und 16Gb DDR4 RAM ausgeführt. Es wurden dabei $k = 7$ Newton-Raphson-Iterationen gewählt, $n = 4$ als durchschnittlichen Bernsteinpolynomgrad sowie Erfolgswahrscheinlichkeit $P(RealBB) = 1 - 10^{-5}$ verwendet. Jede Funktion bekommt dabei maximal ein probabilistisches opakes Prädikat. Benchmarks und Auswertungsskripts sind unter [4] verfügbar.

6.2 Evaluierung

6.2.1 Kosten

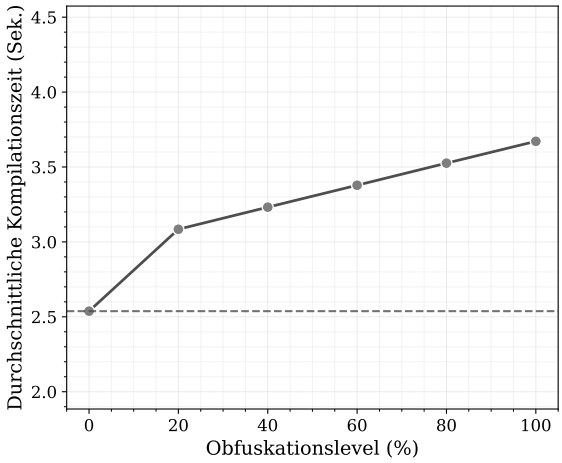

Abbildung 7: Durchschnittliche Kompilationszeit in Abhängigkeit vom Obfuskationslevel

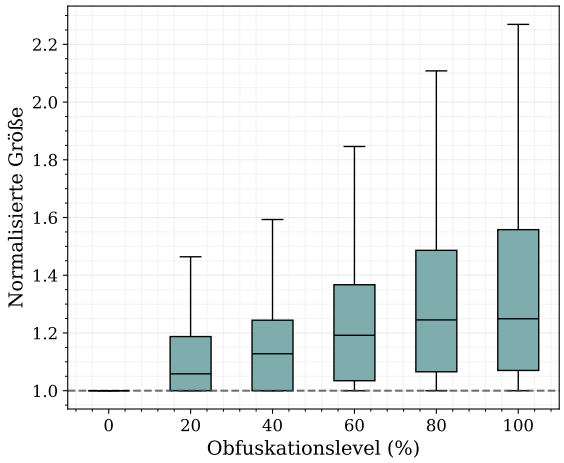

Abbildung 8: Programmgröße in Abhängigkeit vom Obfuskationslevel

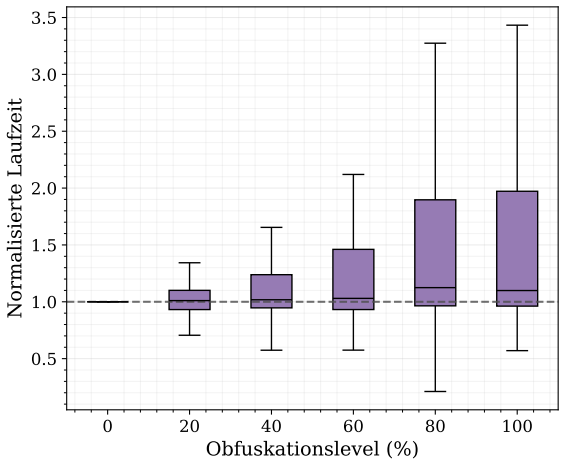

Abbildung 9: Laufzeit in Abhängigkeit vom Obfuskationslevel

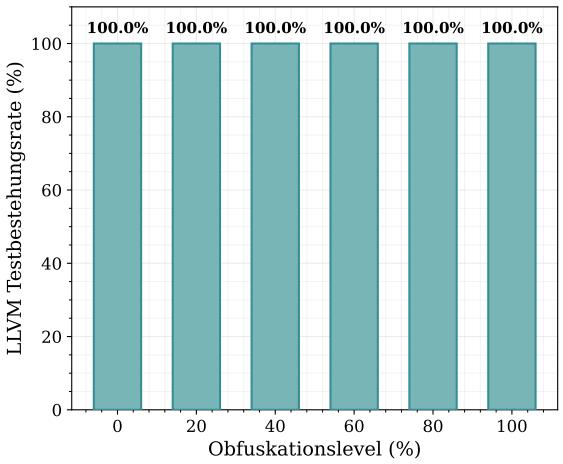

Abbildung 10: Testerfolg in Abhängigkeit vom Obfuskationslevel

Um die Obfuskationskosten zu messen, wurden die durchschnittliche Programmgröße (vgl. Abb. 8), -laufzeit (vgl. Abb. 9), Kompilationszeit (vgl. Abb. 7) und Erfolgsquote (vgl. Abb. 10) von 5 Kompilationen der gesamten

LLVM Test Suite erhoben. Alle Messgrößen wurden test-weise relativ zu Obfuskationslevel 0 normalisiert.

Die Kompilationszeit wächst linear mit der Obfuskationsrate, wie sich Abb. 7 entnehmen lässt.

Wie zu erwarten war zeigt sich ein direkter Zusammenhang zwischen Obfuskationslevel und Code-Größe (vgl. Abb. 8). Während der Median moderat auf ca. 1,25 ansteigt, wächst die Varianz signifikant an. Bei 100% Obfuskation erreichen die Spitzenwerte mehr als das Doppelte ($\sim 2,2$) der Ursprungsgröße. Der Anstieg wird durch das obere Quartil (Q3) getrieben. Während das untere Quartil (Q1) nahe 1,1 verharrt (es gibt also Fälle mit minimalem Overhead), steigt Q3 bei 100% Obfuskation auf ca. 1,6 an. Dass die untere Antenne bis 1,0 reicht ist darauf zurückzuführen, dass sich in 2 Fällen aufgrund fehlender Funktionsparameter keine probabilistischen opaken Prädikate einfügen ließen. Um dies zu korrigieren müsste die Implementierung des Obfuskators zukünftig mit Parameter-Insertion erweitert werden.

Bei der Laufzeit bleibt der Median stabil nahe 1,1, welches auf geringe durchschnittliche Laufzeitkosten hindeutet. Die Streuung (Interquartilsabstand und Minimum/Maximum) nimmt bei höheren Obfuskationsleveln stark zu: 25% der Programmausführungen benötigen die doppelte Laufzeit oder mehr. Die Maxima bei Größe/Laufzeit lassen sich durch die kleineren Programme der Test-Suite erklären, auf welche die Obfuskation eine verhältnismäßig größere Auswirkung hat. Warum manche Programme nach der Obfuskation sogar schneller sind, ist unklar. Das Verhalten wurde bereits bei komerziellen Obfuskationsprodukten beobachtet[6].

Die sich in der Größe oder Laufzeit ergebenden Kosten lassen sich beide durch ein Finetuning der Obfuskatoreinstellungen vermindern, z.B. indem nur bestimmte Funktionen obfuskiert bzw. performance-kritische Funktionen nicht obfuskiert werden.

Trotz der geringen Wahrscheinlichkeit eines Versagens der probabilistischen opaken Prädikate kompilieren und funktionieren auch größere Projekte wie SQLite problemlos/fehlerfrei (vgl. Abb. 10). Insgesamt sind die Kosten je nach Programm/Anwendungsfall und Obfuskationslevel praxisgerecht.

6.2.2 Stärke

Wie in [30] angemerkt wurde, ergibt eine Quantifizierung von Stärke bei der Evaluierung neuer opaker Prädikate wenig Sinn. Relevante Metriken wie die zyklomatische Komplexität (McCabe-Metrik) erhöhen sich beliebig mit der Häufigkeit opaker Prädikate und nicht durch ihre Art.

Auf qualitativer Ebene lässt sich anmerken, dass probabilistische opake Prädikate indirekt auch die Schwachstellen der Analysewerkzeuge heutiger Angreifer angreifen. Konkret sind moderne Dekompilierer wie IDA/Binary Ninja/Ghidra noch nicht in der Lage, bestimmte SIMD (*Single Instruction, Multiple Data*)-Anweisungen zu dekompilieren. Dies lässt sich an Abb. 6c erkennen, wo IDA die unbekannten Anweisungen durch intrinsische Funktionen darstellt anstelle von verständlicherem Pseudo-C. Experimente mit *Decompiler Explorer* [28] zeigen, dass keine der aktuellen *State-of-the-Art*-(SOTA)-Dekompilierer bis auf Ghidra in der Lage ist, die SIMD Anweisungen in Pseudo-C zu dekompilieren. Durch eine Feinjustierung der clang Kommandozeilenoptionen nutzten die probabilistischen opaken Prädikate AVX-512 Anweisungen. Hierdurch konnten die Dekompilierer noch weniger in Pseudo-C dekompilieren. Die Nutzung der *Decompiler Explorer*-Website war hiernach nicht mehr möglich. Diese Limitation bedeutet für Angreifer, dass eine Verständnis der Obfuskation (und somit auch der Programmlogik) mehr Zeit, Aufwand und vor allem ein Verständnis von Assembler erfordert. Die Hürde zum erfolgreichen Reverse-Engineering ist somit erhöht.

6.2.3 Resilienz

Da sich die Resilienz im Falle dieses Projekts nicht quantitativ messen lässt, wird sie qualitativ untersucht. Hierfür wurden folgende Angriffsmethoden auf das einfache Programm aus Abschnitt 5 sowie die 5 Sortieralgorithmen aus [2, 3] angewandt[11]:

- Symbolische Ausführung mit Angr [25] (v.9.2.189),

[11] Gelangt es den symbolischen Ausführungsengines bei diesen einfachen Programmen nicht, den richtigen Basisblock zu erkennen, kann dies bei größeren Programmen auch nicht geschehen.

- Triton [24] (v.1.0.0rc4) und
- Binsec (v.0.11.0) [11]

Wie vermutet waren keine der symbolischen Ausführungsengines in der Lage, die Möglichkeit, dass einer der zwei Pfade nie ausgeführt wird, auszuschließen. Alle drei symbolischen Ausführungsengines erkannten entweder wie hypothetisiert beide Basisblöcke ausgehend vom Prädikat als möglich (Binsec), hängten und konnten nicht innerhalb von 8 Stunden die Constraints des unwahrscheinlichen Pfades extrahieren (Angr) oder konnten gar keine Constraints extrahieren (Triton) (vgl. Tabelle 2).

Engine/Solver	**Version**	**Ergebnis**
Angr (Z3)	9.2.189	✖ keine Constraint-Extraktion (Abbruch nach 8h)
Triton (Z3)	1.0.0rc4	✖ keine Constraints gefunden
Binsec (Bitwuzla)	0.11.0	✖ Beide Basisblöcke sind erreichbar

Tabelle 2: Ergebnisse der symbolischen Ausführung.

Eine probabilistische Analyse über mehrfache Ausführung obfuskierter Programme mit zufälligen Werten [10] erkennt zwar die probabilistischen opaken Prädikate und die korrekten Basisblöcke, markiert aber auch hier fälschlicherweise 20 – 40% regulärer Prädikate der untersuchten Programme als opak. Dies ist damit zu begründen, dass auch viele Prädikate der unobfuskierten Programmlogik besonders häufig gleiche Werte annehmen.

Trotzdem ist eine Deobfuskation theoretisch möglich. Erkennt ein Angreifer ein Prädikat als opak, so muss er nur über dynamischer Ausführung den korrekten Pfad bestimmen. Die Resilienz der Obfuskation basiert somit auf ihrer Tarnung. Wie Abschnitt 6.2.4 zeigt ist eine wiederholte automatisierte Erkennung bei einer Kombination mit anderen Obfuskationsmethoden allerdings schwierig.

Zudem ist eine Anwendung von an die Obfuskationsmethode angepassten Heuristiken möglich. Ein Angreifer könnte z.B. alle Prädikate mit symbolischen Ausführungsengine ausführen und die benötigte Zeit für die Extraktion der Constraints beider Pfade messen. Dauert dies für einen Pfad länger als eine festgelegte konstante Zeit, so handelt es sich mit hoher Wahrscheinlichkeit um ein probabilistisch opakes Prädikat. Trotzdem hat die Methode den gleichen Nachteil wie die probabilistische Analyse – ihre Korrektheit ist nie garantiert.

6.2.4 Tarnung

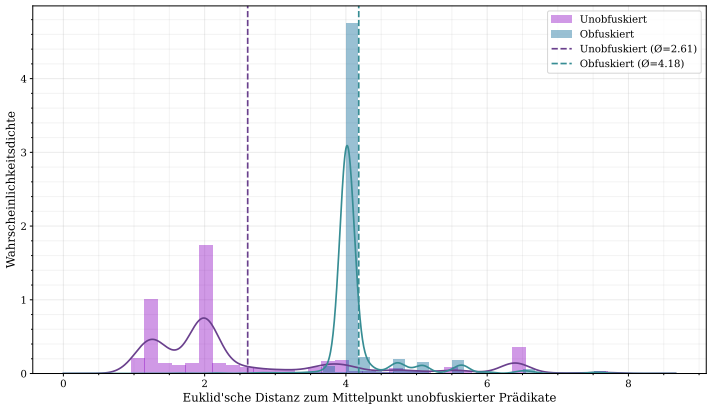

Abbildung 11: Histogramm und KDE der Wahrscheinlichkeitsdichte der euklidischen Distanz zum Mittelpunkt unobfuskierter Prädikate vor/nach der Obfuskation

Um die Tarnung quantitativ zu messen wurde wie bei [30] vorgegangen. Hierfür wurden $n = 4500$ jeweils unobfuskierte sowie obfuskierte Prädikate zufällig gewählt. Gemäß Tabelle 3 wurden die letzten zehn Anweisungen vor der bedingten Sprungsanweisung (inkl. der Sprunganweisung selbst) kategorisiert. Hierdurch kann jedes

Kategorie	Anweisungen
Arithmetik	`imul, inc, sub, add, idiv, divsd, sbb, subpd, addpd, mulpd, divpd, maxpd, minpd, sqrtpd`
Logik	`and, sar, xor, test, shr, shl, or, xorps, xorpd, orpd, andpd`
Datenübertragung	`movaps, movsd, movabs, movzx, mov, movss, movsx, movsxd, stosd, movapd, movupd, unpcklpd, unpckhpd, movhpd, movlpd`
Datenbreite/-konversion	`cvtss2sd, cvtsi2sd, cvtsd2ss, cqo, cdq, cvttsd2si`
Zeigerarithmetik	`lea`
Vergleich	`cmp, ucomisd, comisd`
Sprünge	`jle, jne, jge, jae, jl, je, jg, jp, ja, jbe, jno, jmp`
Stapel	`pop, push, call, ret`
Boolesch	`setge, setne, setg, seta, setb, setl, sete`
Sonstige	`nop`

Tabelle 3: 74 Assembler-Anweisungen in 10 Kategorien sortiert.

Prädikat i als 10-dimensionaler Vektor $\vec{x}_i$ dargestellt werden, wobei jede Dimension der absoluten Häufigkeit einer Anweisungskategorie in den letzten 10 Anweisungen vor der Sprunganweisung entspricht. Um den Unterschied zwischen obfuskierten und unobfuskierten Prädikaten zu messen wurde zuerst der Mittelpunkt $\vec{m} = \frac{\sum_{i=1}^{n} \vec{x}_i}{n}$ aller unobfuskierten Prädikate berechnet. Die durchschnittliche Euklid'sche Distanz der obfuskierten Prädikate zum Mittelpunkt $\vec{m}$ entspricht ihrer Ähnlichkeit und somit auch Tarnung. Je geringer der Unterschied ist, desto höher die Tarnung.

Wie Abb. 11 zu entnehmen ist, liegt die durchschnittliche Euklid'sche Distanz probabilistischer opaker Prädikate zu $\vec{m}$ bei $4,18$. Die durchschnittliche Distanz regulärer Prädikate zu $\vec{m}$ liegt bei $2,61$. Die Distanz ist vor allem bei Programmen mit nur wenigen Gleitkommazahlberechnungen hoch. Die Tarnung lässt sich hier erhöhen durch eine Kombination mit anderen Obfuskationsmethoden, z.B. dem Einfügen von „Dead Code"/„Junk Code" mit vielen Gleitkommazahloperationen.

Eine automatisierte Erkennung mittels Pattern-Matching erwies sich empirisch als schwierig. Es konnten keine Patterns erstellt werden, welche alle probabilistisch opaken Prädikate abdeckten. Eine weitere Härtung ist über die Anwendung von *Mixed Boolean Arithmetic* (MBA) Obfuskation auf das gesamte Programm möglich. Eine automatisierte Erkennung wird hier über die variablen Ausdrücke verhindert. Eine manuelle Erkennung der opaken Prädikate durch Angreifer ist behindert durch die Ununterscheidbarkeit der MBA Ausdrücke in den probabilistischen opaken Prädikaten vom Rest der obfuskierten Ausdrücke des Programms.

7 Fazit

In der vorliegenden Arbeit wurde das neue Softwareobfuskationskonzept der probabilistischen opaken Prädikate eingeführt und in einer Implementierung in Form eines LLVM-Passes ein Machbarkeitsnachweis geliefert. Empirische Untersuchungen bestätigen, dass die vorgestellte Idee hinsichtlich ihrer Kosten, Resilienz, Stärke und Tarnung je nach Anwendungsfall und Kombination mit anderen Obfuskationsmethoden praxisgerecht ist. Dabei konnte nebenbei ein undokumentierter Angriffsvektor erkannt und eine Lösung dafür geboten werden.

Die Resilient probabilistischer opaker Prädikate gegen symbolische Ausführung wurde theoretisch und praktisch begründet. Praktische Untersuchungen zeigen eine Resilienz gegen weitere Angriffsmethoden, wie einer probabilistischen Analyse sowie Pattern-Matching. Die Forschungsfrage ist somit positiv beantwortet.

Offen bleibt die Resilienz gegenüber Machine Learning basierten Deobfuskationattacken wie [27]. Auch mögliche Symbiosen mit verschiedenen Obfuskationsmethoden sollten weiter untersucht werden. Zukünftige Arbeiten sollten außerdem weitere Methoden zur Generierung probabilistischer opaker Prädikate untersuchen. So könnten z.B. Wahrscheinlichkeitsverteilungen mit Extremum verwendet werden, um bestimmte Werte besonders häufig vorkommen zu lassen und so opak *wahr* (1) oder *falsch* (0) darzustellen. Dies konnte aufgrund zeitlicher Be-

schränkungen in dieser Arbeit nicht behandelt werden. Gegenstand einer weiterführenden Auseinandersetzung sollte zudem die Anwendung der hier dargelegten Prinzipien der probabilistischen Obfuskation gegen andere Deobfuskationsmethoden sein. So könnte z.B. die Anwendung auf Ausdrücke Programmsynthesemethoden [5, 20] verhindern.

Danksagung

Ich danke meinem schulischen Projektbetreuer, Herrn Dr. Arndt Latußeck für die wichtige Hinweisen zur Gliederung, Vorschläge zum Vorgehen bei der Evaluierung sowie für fachliche Vorschläge zur vorliegenden Arbeit.

Zudem danke ich meiner Mutter, Claudia Baumgartner-Bardubitzki für ihre sprachliche Überprüfung.

Des Weiteren danke ich allen Entwicklern verschiedener Open-Source Softwarebibliotheken und -Anwendungen. Ohne diese wäre eine Ausarbeitung in der begrenzten Zeit und aktueller Form unmöglich gewesen.

Literaturverzeichnis

[1] Baldoni, Roberto u. a. *A Survey of Symbolic Execution Techniques*. Mai 2018. DOI: 10.48550/arXiv.1610.00502. eprint: 1610.00502 (cs). (Besucht am 27. 07. 2025).

[2] Banescu, Sebastian u. a. "Code Obfuscation against Symbolic Execution Attacks". In: *Proceedings of the 32nd Annual Conference on Computer Security Applications*. Los Angeles California USA: ACM, Dez. 2016, S. 189–200. ISBN: 978-1-4503-4771-6. DOI: 10.1145/2991079.2991114. (Besucht am 19. 12. 2025).

[3] Banescu, Sebastian u. a. *Quellcode für Obfuscation Benchmarks*. https://github.com/tum-i4/obfuscation-benchmarks. 2025. (Besucht am 29. 12. 2025).

[4] Baumgartner, Paul. *Quelltext für die Implementierung von POP*. https://github.com/sariaki/JuFo-2026. 2025. (Besucht am 29. 12. 2025).

[5] Blazytko, Tim u. a. "Syntia: Synthesizing the semantics of obfuscated code". In: *26th USENIX Security Symposium (USENIX Security 17)*. 2017, S. 643–659.

[6] CodeDefender. *Metrics: Performance metrics of CodeDefender vs Themida vs VMProtect*. GitHub repository, inkl. Benchmarks und Rohdaten. CodeDefender. URL: https://github.com/codedefender-io/metrics (besucht am 12. 01. 2026).

[7] Collberg, Christian. *The Tigress C Obfuscator*. https://tigress.wtf/. 2025. (Besucht am 14. 10. 2025).

[8] Collberg, Christian, Thomborson, Clark und Low, Douglas. "A Taxonomy of Obfuscating Transformations". In: *http://www.cs.auckland.ac.nz/staff-cgi-bin/mjd/csTRcgi.pl?serial* (Jan. 1997).

[9] Collberg, Christian, Thomborson, Clark und Low, Douglas. "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).

[10] Dalla Preda, Mila u. a. "Opaque Predicates Detection by Abstract Interpretation". In: *Algebraic Methodology and Software Technology*. Hrsg. von Hutchison, David u. a. Bd. 4019. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, S. 81–95. ISBN: 978-3-540-35633-2 978-3-540-35636-3. DOI: 10.1007/11784180_9. (Besucht am 01. 01. 2026).

[11] Daniel, L., Bardin, S. und Rezk, T. "Binsec/Rel: Efficient Relational Symbolic Execution for Constant-Time at Binary-Level". In: *2020 IEEE Symposium on Security and Privacy (SP)*. Los Alamitos, CA, USA: IEEE Computer Society, Mai 2020, S. 1021–1038. DOI: 10.1109/SP40000.2020.00074. URL: https://doi.ieeecomputersociety.org/10.1109/SP40000.2020.00074.

[12] Defense Advanced Research Projects Agency. *SafeWare*. 2025. URL: https://www.darpa.mil/research/programs/safeware (besucht am 29. 12. 2025).

[13] Georgii, Hans-Otto. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 5. Auflage. De Gruyter Studium. Berlin ; Boston: De Gruyter, 2015. 438 S. ISBN: 978-3-11-035969-5.

[14] Junod, Pascal 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 Wyseur, Brecht. IEEE, 2015, S. 3–9. DOI: 10.1109/SPRO.2015.10.

[15] Lattner, Chris und Adve, Vikram. "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.

[16] Lin, Hong u. a. “Branch Obfuscation Using Binary Code Side Effects”. In: *Proceedings of the International Conference on Computer, Networks and Communication Engineering (ICCNCE 2013)*. The International Conference on Computer, Networks and Communication Engineering (ICCNCE 2013). China: Atlantis Press, 2013. DOI: 10.2991/iccnce.2013.37. URL: http://www.atlantis-press.com/php/paper-details.php?id=6493 (besucht am 13. 07. 2025).

[17] LLVM Project. *The LLVM Compiler Infrastructure*. https://llvm.org/. Version 20.1.4. 2003–2025. (Besucht am 18. 08. 2025).

[18] LLVM Test Suite Contributer. *llvm-test-suite*. https://github.com/llvm/llvm-test-suite. 2025.

[19] Lorentz, G. G. *Bernstein Polynomials*. University of Toronto Press / Springer (reprint/excerpt available), 1953.

[20] Menguy, Grégoire u. a. “Search-Based Local Black-Box Deobfuscation: Understand, Improve and Mitigate”. In: *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. CCS ’21: 2021 ACM SIGSAC Conference on Computer and Communications Security. Virtual Event Republic of Korea: ACM, 12. Nov. 2021, S. 2513–2525. ISBN: 978-1-4503-8454-4. DOI: 10.1145/3460120.3485250. URL: https://dl.acm.org/doi/10.1145/3460120.3485250 (besucht am 09. 01. 2026).

[21] Ogiso, Toshio u. a. “Software obfuscation on a theoretical basis and its implementation”. In: *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences* 86.1 (2003), S. 176–186.

[22] Ollivier, Mathilde u. a. “How to Kill Symbolic Deobfuscation for Free (or: Unleashing the Potential of Path-Oriented Protections)”. In: *Proceedings of the 35th Annual Computer Security Applications Conference*. San Juan Puerto Rico USA: ACM, Dez. 2019, S. 177–189. ISBN: 978-1-4503-7628-0. DOI: 10.1145/3359789.3359812. (Besucht am 22. 12. 2025).

[23] Reichenwallner, Benjamin und Meerwald-Stadler, Peter. “Simplification of General Mixed Boolean-Arithmetic Expressions: GAMBA”. In: *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. Juli 2023, S. 427–438. DOI: 10.1109/EuroSPW59978.2023.00053. arXiv: 2305.06763 [cs]. URL: http://arxiv.org/abs/2305.06763 (besucht am 14. 06. 2025).

[24] Saudel, Florent und Salwan, Jonathan. “Triton: A Dynamic Symbolic Execution Framework”. In: *Symposium sur la sécurité des technologies de l’information et des communications*. SSTIC. Rennes, France, Juni 2015, S. 31–54.

[25] Shoshitaishvili, Yan u. a. “SoK: (State of) The Art of War: Offensive Techniques in Binary Analysis”. In: *IEEE Symposium on Security and Privacy*. 2016.

[26] Sutter, Bjorn De u. a. *Evaluation Methodologies in Software Protection Research*. Apr. 2024. DOI: 10.48550/arXiv.2307.07300. arXiv: 2307.07300 [cs]. (Besucht am 19. 12. 2025).

[27] Tofighi-Shirazi, Ramtine 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.

[28] Vector 35. *Quelltext von Decompiler Explorer*. https://github.com/decompiler-explorer/decompiler-explorer. 2025. (Besucht am 29. 12. 2025).

[29] Wang, Zhi u. a. “Linear Obfuscation to Combat Symbolic Execution”. In: *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: http://link.springer.com/10.1007/978-3-642-23822-2_12 (besucht am 22. 07. 2025).

[30] Xu, Hui 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 Networks (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).

[31] Zobernig, Lukas, Galbraith, Steven D. und Russello, Giovanni. “When Are Opaque Predicates Useful?” In: *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). Rotorua, New Zealand: IEEE, Aug. 2019, S. 168–175. DOI: 10.1109/trustcom/bigdatase.2019.00031. URL: https://ieeexplore.ieee.org/document/8887369/ (besucht am 17. 07. 2025).

Abbildungsverzeichnis