

Reverse Engineering 101

Dozent: Prof. Dr. Michael Eichberg
Kontakt: michael.eichberg@dhbw-mannheim.de
Version: 1.0

Folien:

- <https://delors.github.io/sec-reversing101/folien.de.rst.html>
- <https://delors.github.io/sec-reversing101/folien.de.rst.html.pdf>

Fehler auf Folien melden:
<https://github.com/Delors/delors.github.io/issues>

Vorerfahrungen?

- Wer hat schon einmal Software or Hardware Reverse Engineering betrieben?
- Wer kennt Java Bytecode?
- Wer hat Erfahrung mit Python?

Reverse Engineering

Reverse Engineering ist die Analyse von Systemen mit dem Ziel, ihren Aufbau und ihre Funktionsweise zu verstehen.

Typische Anwendungsfälle:

- die Rekonstruktion (von Teilen) des Quellcodes von Programmen, die nur als Binärabbild vorliegen.
- die Analyse von Kommunikationsprotokollen proprietärer Software

Vom Reverse Engineering ist das **Reengineering** zu unterscheiden. Im Fall von letzteren geht es „nur“ darum die Funktionalität eines bestehenden Systems mit neuen Techniken wiederherzustellen.

Zweck von Reverse Engineering

- Herstellung von Interoperabilität
- Untersuchung auf Schwachstellen
- Untersuchung auf Copyrightverletzungen
- Untersuchung auf Backdoors
- Analyse von Viren, Würmern etc.
- Umgehung von ungerechtfertigten(?) Schutzmaßnahmen (z.B. bei Malware)

CVE-2024-3094 - liblzma Backdoor in OpenSSH[1][2]

Ziel	Verbreitung des Schadcode?	Bewertung
<p>Das Verhalten von SSH bei der Authentifikation so zu verändern, dass es dem Angreifer Zugang zum System erlaubt.</p> <p>Zur Absicherung der Backdoor ist diese über ein Zertifikat abgesichert.</p>	<p>Die Bibliothek <code>liblzma</code> wurde so angepasst, dass diese eine Backdoor in SSH einbaut.</p> <p>Der Schadcode ist nur in den Tarballs zu finden - nicht im SourceCode im GIT. Der eigentliche Schadcode wurde versteckt in <code>Testfixtures</code>.</p> <p>Der Code wurde so entworfen, dass bekannte Werkzeuge (<code>Valgrind</code>) keine Probleme erkennen.</p> <p>Die Bibliothek wurde nur in bestimmten Situationen von OpenSSH verwendet.</p>	<p><i>CVSS Base Score:</i> 10.0 (kritisch)</p> <p><i>Entstandener Schaden:</i> vermutlich gering, da (gerade noch) keine offiziellen Releases (von Debian, Ubuntu, etc.) betroffen waren.</p> <p>Dem Angriff ging ein sehr langer Social Engineering Angriff voraus, weswegen mit höherer Wahrscheinlichkeit ein „State-sponsored Actor“ dahintersteckt.</p>

[1] InnoQ Podcast

[2] SSH Blob

Backdoor in 16 D-Link Routern[3]

- Angreifer können aus dem lokalen Netzwerk heraus den Telnet-Dienst betroffener D-Link-Router durch Angabe einer bestimmten Ziel URL aktivieren.
- Die Admin-Zugangsdaten sind in der Firmware hinterlegt.
- Vermutlich ursprünglich für werksseitige Tests.
- *CVSS Base Score*: 8.8 (hoch)

[3] [Golem.de](#)

Reverse Engineering - grundlegende Schritte

- 01** Informationsgewinnung zur Gewinnung aller relevanten Informationen über das Produkt.
- 02** Modellierung mit dem Ziel der (Wieder-)Gewinnung eines (abstrakten) Modells der relevanten Funktionalität.
- 03** Überprüfung (🇺🇸 *review*) des Modells auf seine Richtigkeit und Vollständigkeit.

Informationsgewinnung - Beispiel

Gegeben sei eine App zum Ver- und Entschlüsseln von Dateien sowie ein paar verschlüsselte Dateien. Mögliche erste Schritte vor der Analyse von Binärcode:

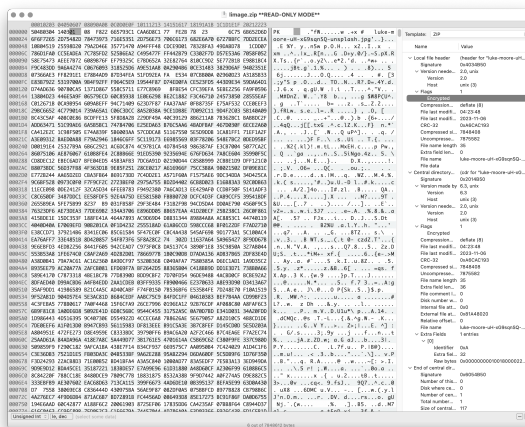
- Die ausführbare Datei ggf. mit **file** (oder sogar mit **binwalk**) überprüfen (z.B. wie wurde die Datei kompiliert und für welches Betriebssystem und Architektur)

Beispiel:

```
$ file /usr/bin/openssl
/usr/bin/openssl: Mach-O universal binary with 2 archi...
/usr/bin/openssl (for architecture x86_64): Mach-O 64-bit
/usr/bin/openssl (for architecture arm64e): Mach-O 64-bit
```

1

- Die Dateien mit einem (guten) Hexeditor auf Auffälligkeiten untersuchen.



2

Die Datei auf bekannte Viren und Malware überprüfen.

3

- Eine Datei mit einem bekannten Inhalt verschlüsseln und danach vergleichen. Ist die Datei gleich groß?

Falls ja, dann werden keine Metainformationen gespeichert und das Passwort kann (ggf.) nicht (leicht) verifiziert werden.

(Es kann zumindest nicht direkt in der Datei gespeichert sein.)

4

- Eine Datei mit verschiedenen Passwörtern verschlüsseln.

Sind die Dateien gleich?

Falls ja, dann wäre die Verschlüsselung komplett nutzlos und es gilt nur noch den konstanten Schlüssel zu finden.

Gibt es Gemeinsamkeiten?

Falls ja, dann wäre es möglich, dass das Passwort (gehasht) in der Datei gespeichert wird.

5

- Eine Datei mit einem wohldefinierten Muster verschlüsseln, um ggf. den „Mode of Operation“ (insbesondere ECB) zu identifizieren.

6

- Mehrere verschiedene Dateien mit dem gleichen Passwort verschlüsseln

Gibt es Gemeinsamkeiten?

Falls ja, dann wäre es möglich, dass die entsprechenden Teile direkt vom Passwort abgeleitet werden/damit verschlüsselt werden.

7

- ...

8

- Reverse Engineering der App durchführen.

9

Rechtliche Aspekte des Reverse Engineering

Vorsicht!

Die Gesetzgebungen unterscheiden sich von Land zu Land teils signifikant.

- Die Rechtslage hat sich in Deutschland mehrfach geändert.
- Umgehung von Kopierschutzmechanismen ist im Allgemeinen verboten.
- Lizenz verbietet das Reverse Engineering häufig!

Warnung

Bevor Sie Reverse Engineering von Systemen betreiben, erkundigen sie sich erst über mögliche rechtliche Konsequenzen.

1. SOFTWARE REVERSE ENGINEERING

Prof. Dr. Michael Eichberg

Ansätze

statische Analyse: Studieren des Programms ohne es auszuführen; typischerweise mittels eines Disassemblers oder eines Decompilers.

dynamische Analyse:

Ausführen des Programms; typischerweise unter Verwendung eines Debuggers oder eines instrumentations Frameworks (z.B. *Frida*).

hybride Analyse: Kombination aus statischer und dynamischer Analyse.

Ansätze wie *Unicorn*, welches auf *QEmu* aufbaut, erlaubt zum Beispiel die Ausführung von (Teilen von) Binärcode auf einer anderen Architektur als der des Hosts.

Ein Beispiel wäre die Ausführung einer Methode, die im Code verschlüsselte hinterlegte Strings entschlüsselt (*deobfuscation*), um die Analyse zu vereinfachen.

Ggf. müssen für Teile des Codes, die die Hostfunktionalität nutzen, Stubs/Mocks bereitgestellt werden.

Disassembler

Überführt (maschinenlesbaren) Binärcode in Assemblercode

Kommandozeilenwerkzeuge (exemplarisch):

- `objdump -d`
- `gdb`
- `radare`
- `javap` (für Java)

Hinweis

Für einfache Programme ist es häufig möglich direkt den gesamten Assemblercode mittels der entsprechenden Werkzeuge zu erhalten. Im Falle komplexer Binärdateien (z.B. im ELF (Linux) und PE (Windows) Format) gilt dies nicht und erfordert ggf. manuelle Unterstützung zum Beispiel durch das Markieren von Methodenanfängen.

Im Fall von Java `.class` ist die Disassemblierung immer möglich.

Decompiler

Überführt (maschinenlesbare) Binärcode *bestmöglich* in Hochsprache (meist C ähnlich oder Java). Eine *kleine* Auswahl von verfügbaren Werkzeugen:

- Hex-Rays IDAPro (kommerziell)
- Ghidra (unterstützt fast jede Plattform; die Ergebnisse sind sehr unterschiedlich.)
- JadX (Androids `.dex` Format)
- CFR (Java `.class` Dateien)
- IntelliJ

Hinweis

Decompiler sind generell sehr hilfreich, aber gleichzeitig auch sehr fehlerbehaftet. Vieles, was im Binärcode möglich ist, hat auf der Ebene des Sourcecodes keine Entsprechung.

Zum Beispiel unterstützt Java Bytecode beliebige Sprünge. Solcher Code wird aber durch normale Programme, die z. B. in Java, Kotlin, Scala oder Clojure geschrieben wurden, nicht erzeugt. Decompiler kommen mit solchem Code in der Regel nicht (gut) zurecht.

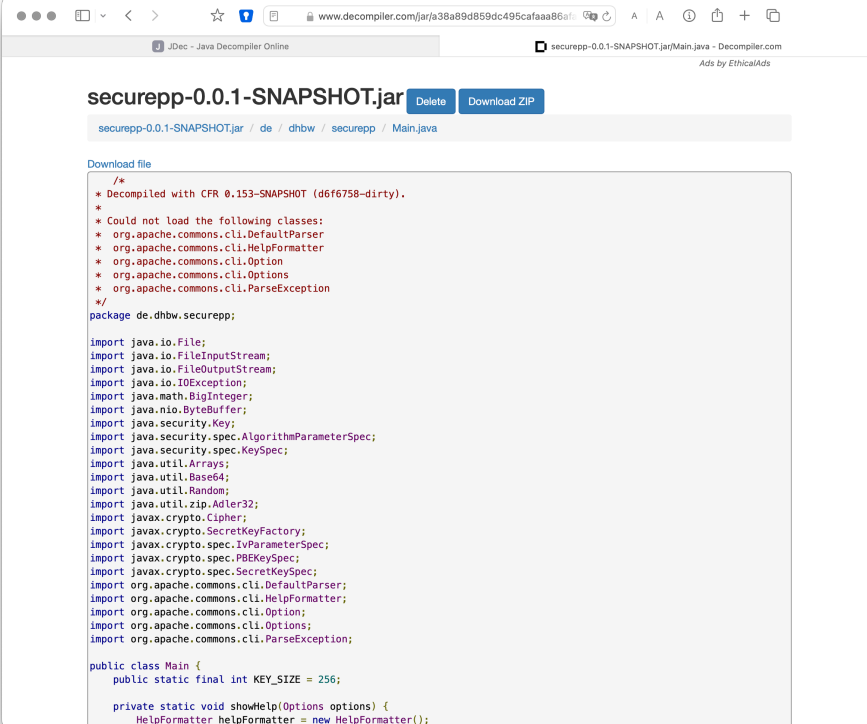
13

Mittels Decompiler ist es ggf. möglich Code, der zum Beispiel ursprünglich in Kotlin oder Scala geschrieben und für die JVM kompiliert wurde, als Java Code zurückzubekommen.

Die Ergebnisse sind für Analysezwecke zwar häufig ausreichend gut – von funktionierendem Code jedoch ggf. (sehr) weit entfernt.

decompiler.com unterstützt eine große Anzahl ausführbaren Dateien.

cfr Decompiler



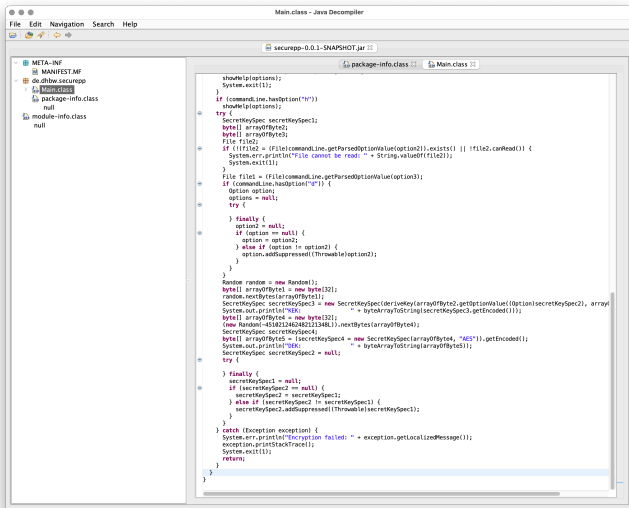
```
/*
 * Decompiled with CFR 0.153-SNAPSHOT (d6f6758-dirty).
 *
 * Could not load the following classes:
 * org.apache.commons.cli.DefaultParser
 * org.apache.commons.cli.HelpFormatter
 * org.apache.commons.cli.Option
 * org.apache.commons.cli.Options
 * org.apache.commons.cli.ParseException
 */
package de.dhbw.securepp;

import java.io.File;
import java.io.FileInputStream;
import java.io.FileOutputStream;
import java.io.IOException;
import java.math.BigInteger;
import java.nio.ByteBuffer;
import java.security.Key;
import java.security.spec.AlgorithmParameterSpec;
import java.security.spec.KeySpec;
import java.util.Arrays;
import java.util.Base64;
import java.util.Random;
import java.util.zip.Adler32;
import javax.crypto.Cipher;
import javax.crypto.SecretKeyFactory;
import javax.crypto.spec.IvParameterSpec;
import javax.crypto.spec.PBEKeySpec;
import javax.crypto.spec.SecretKeySpec;
import org.apache.commons.cli.DefaultParser;
import org.apache.commons.cli.HelpFormatter;
import org.apache.commons.cli.Option;
import org.apache.commons.cli.Options;
import org.apache.commons.cli.ParseException;

public class Main {
    public static final int KEY_SIZE = 256;

    private static void showHelp(Options options) {
        HelpFormatter helpFormatter = new HelpFormatter();
    }
}
```

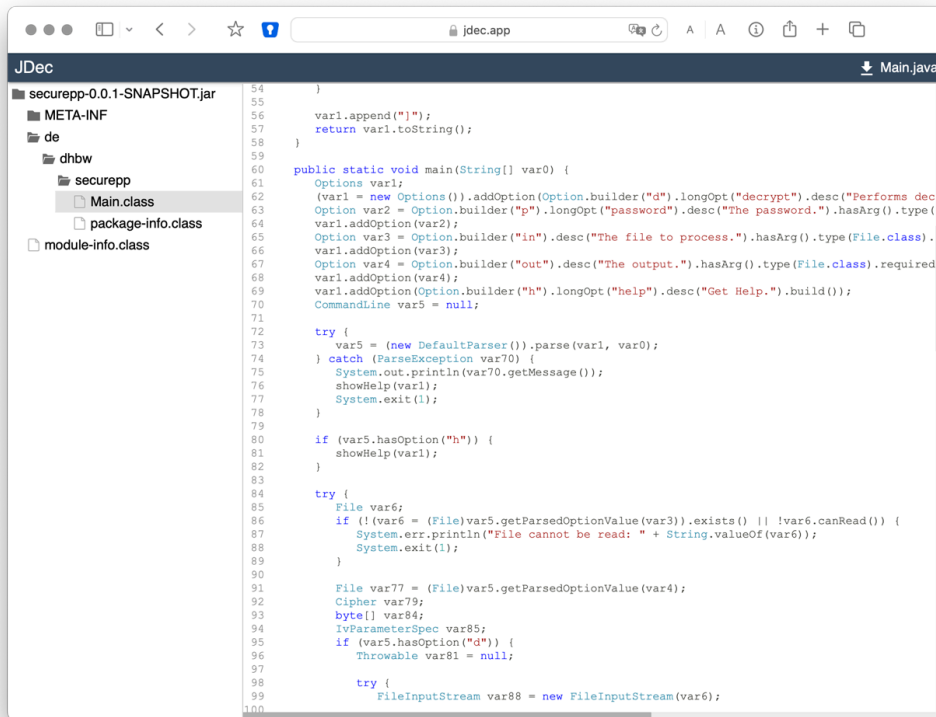
JD Decompiler



```
SecretKeySpec secretKeySpec2 = null;
try {
} finally {
    secretKeySpec1 = null;
    if (secretKeySpec2 == null) {
        secretKeySpec2 = secretKeySpec1;
    } else if (secretKeySpec2 != secretKeySpec1) {
        secretKeySpec2.addSuppressed((Throwable)secretKeySpec1);
    }
}
```

Beispiel fehlgeschlagener Dekompilierung

JDec Decompiler



The screenshot shows the JDec application window with a file explorer on the left and a code editor on the right. The file explorer shows the project structure for 'securepp-0.0.1-SNAPSHOT.jar', including 'META-INF', 'de', 'dhbw', 'securepp', 'Main.class', 'package-info.class', and 'module-info.class'. The code editor displays the decompiled Java code for 'Main.class', starting with a package declaration and a main method that uses the 'Options' class for command-line argument parsing.

```
54     }
55
56     var1.append("");
57     return var1.toString();
58 }
59
60 public static void main(String[] var0) {
61     Options var1;
62     (var1 = new Options()).addOption(Option.builder("d").longOpt("decrypt").desc("Performs dec
63 Option var2 = Option.builder("p").longOpt("password").desc("The password.").hasArg().type(
64 var1.addOption(var2);
65 Option var3 = Option.builder("in").desc("The file to process.").hasArg().type(File.class).
66 var1.addOption(var3);
67 Option var4 = Option.builder("out").desc("The output.").hasArg().type(File.class).required
68 var1.addOption(var4);
69 var1.addOption(Option.builder("h").longOpt("help").desc("Get Help.").build());
70 CommandLine var5 = null;
71
72 try {
73     var5 = (new DefaultParser()).parse(var1, var0);
74 } catch (ParseException var70) {
75     System.out.println(var70.getMessage());
76     showHelp(var1);
77     System.exit(1);
78 }
79
80 if (var5.hasOption("h")) {
81     showHelp(var1);
82 }
83
84 try {
85     File var6;
86     if (!(var6 = (File)var5.getParsedOptionValue(var3)).exists() || !var6.canRead()) {
87         System.err.println("File cannot be read: " + String.valueOf(var6));
88         System.exit(1);
89     }
90
91     File var77 = (File)var5.getParsedOptionValue(var4);
92     Cipher var79;
93     byte[] var84;
94     IvParameterSpec var85;
95     if (var5.hasOption("d")) {
96         Throwable var81 = null;
97
98         try {
99             FileInputStream var88 = new FileInputStream(var6);
100
```

Debugger

Dient der schrittweisen Ausführung des zu analysierenden Codes oder Hardware; ermöglichen zum Beispiel Speicherinspektion und Manipulation.

- gdb
- lldb
- x64dbg (Windows, Open-Source)
- jdb (Java Debugger)

Hardware Debugger

Für das Debuggen von Hardware gibt es entsprechende Werkzeuge, z.B. [Lauterbach Hardware Debugger](#) (kommerziell und sehr teuer).

Mittels solcher Werkzeuge ist es möglich die Ausführung von Hardware Schritt für Schritt (🇺🇸 *single step mode*) zu verfolgen und den Zustand der Hardware (Speicher und Register) zu inspizieren. Dies erfordert jedoch häufig eine JTAG Schnittstelle oder etwas vergleichbares.

2. ERSCHWERUNG DES REVERSE ENGINEERING

Prof. Dr. Michael Eichberg

Obfuscation (🇩🇪 Verschleierung)

- Techniken, die dazu dienen das Reverse Engineering zu erschweren.
- Häufig eingesetzt ...
 - von Malware
 - Adware (im Kontext von Android ein häufig beobachtetes Phänomen)
 - zum Schutz geistigen Eigentums
 - für DRM / Durchsetzung von Kopierrechten
 - zur Prävention von „Cheating“ (insbesondere im Umfeld von Online Games)
 - Wenn das Programm als Source Code vertrieben wird bzw. vertrieben werden muss (JavaScript)
- Arbeiten auf Quellcode oder Maschinencode Ebene
- Grenze zwischen *Code Minimization*, *Code Optimization* und *Code Obfuscation* ist fließend.
- Mögliche Werkzeuge (ohne Wertung der Qualität/Effektivität):
 - [Java] Proguard / Dexguard
 - [C/C++] **Star Force**

19

Gerade im Umfeld von klassischen *Binaries* für Windows, Mac und Linux erhöhen Compiler Optimierungen, z.B. von C/C++ und Rust Compilern (-O2 / -O3), bereits den Aufwand, der notwendig ist den Code zu verstehen, erheblich.

Hinweis

Einen ambitionierten und entsprechend ausgestatteten Angreifer wird **Code Obfuscation** bremsen, aber sicher nicht vollständig ausbremsen und das Vorhaben verteilen.

Obfuscation - Techniken (Auszug)

- entfernen aller Debug-Informationen
- Das Kürzen aller möglichen Namen (insbesondere Methoden und Klassennamen).
- Das Verschleiern von Konstanten durch den Einsatz vermeintlich komplexer Berechnungen zu deren Initialisierung.

```
~(((int)Math.PI) ^ Integer.MAX_VALUE >> 16)+Short.MAX_VALUE  
= 2
```

- Die Verwendung von Unicode Codepoints für Strings oder die Verschleierung von Strings mittels rot13 Verschlüsselung.

```
/* ??? */ printf("\x48""e\x154\x6F"" \127o\x72""\144!");  
/* = */ printf("Hello World!");
```

- Das Umstellen von Instruktionen, um das Dekompilieren zu erschweren.
- Das Hinzufügen von totem Code.
- Den relevanten Teil der Anwendung komprimieren und verschlüsseln und erst bei Verwendung entpacken und entschlüsseln.
- ...

Obfuscation auf Source Code Ebene: [International Obfuscated C Code Contest](#)

Umstellen von Instruktionen

Das Umstellen von Instruktionen erschwert die Analyse, da viele Werkzeuge zum Dekompilieren auf die Erkennung von bestimmten Mustern im Code angewiesen sind und ansonsten nur sehr generischen (Spagetti Code) oder gar unsinnigen Code zurückgeben.

Verschleierung von Strings

Das Verschleiern von Strings kann insbesondere das Reversen von Binärcode erschweren, da ein Angreifer häufig „nur“ an einer ganz bestimmten Funktionalität interessiert ist und dann Strings ggf. einen sehr guten Einstiegspunkt für die weitergehende Analyse bieten.

Stellen Sie sich eine komplexe Java Anwendung vor, in der alle Namen von Klassen, Methoden und Attributen durch einzelne oder kurze Sequenzen von Buchstaben ersetzt wurden und sie suchen danach wie von der Anwendung Passworte verarbeitet werden. Handelt es sich um eine GUI Anwendung, dann wäre zum Beispiel die Suche nach Text, der in den Dialogen vorkommt (z.B. "Password") z.B. ein sehr guter Einstiegspunkt.

3. EINE SEHR KURZ EINFÜHRUNG IN JAVA BYTECODE

Prof. Dr. Michael Eichberg

Die Java Virtual Machine

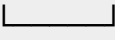
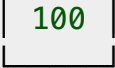
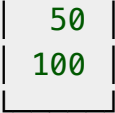

- **Java Bytecode** ist die Sprache, in der Java (oder Scala, Kotlin, ...) Programme auf der Java Virtual Machine (JVM) [4] ausgeführt werden.
- In den meisten Fällen arbeiten Java Decompiler so gut, dass ein tiefgehendes Verständnis von Java Bytecode selten notwendig ist.
- Java Bytecode kann — muss aber nicht — interpretiert werden. (Z.B. können „virtuelle Methodenaufrufe“ in Java schneller sein als in C++.)

[4] Java Bytecode Spezifikation

Java Bytecode - stackbasierte virtuelle Maschine

Die JVM ist eine stackbasierte virtuelle Maschine.

Die getypten Operanden eines Befehls werden auf einem Stack abgelegt und die Operationen arbeiten auf den obersten Elementen des Stacks. Jeder Thread hat seinen eigenen Stack.

Instruktionen	Veränderung des Stacks
nop	
bipush 100 → int	
bipush 50 → int	
iadd ← 2 × int → int	

Eine Methode muss einen Stack begrenzter Höhe aufweisen. Code, für den die Stackhöhe nicht berechenbar ist, wird vom Compiler abgelehnt. (Zum Beispiel ein **bipush** in einer Endlosschleife.) Die benötigte Höhe des Stacks wird vom Compiler berechnet und von der JVM überprüft.

Java Bytecode - Methodenaufrufe und lokale Variablen

- Die Java Virtual Machine verwendet lokale Variablen zur Übergabe von Parametern beim Methodenaufruf.
- Beim Aufruf von *Klassenmethoden* (**static**) werden alle Parameter in aufeinanderfolgenden lokalen Variablen übergeben, beginnend mit der lokalen Variable 0. D.h. in der aufrufenden Methode werden die Parameter vom Stack geholt und in lokalen Variablen gespeichert.
- Beim Aufruf von *Instanzmethoden* wird die lokale Variable 0 dazu verwendet, um die Referenz (**this**) auf das Objekt zu übergeben, auf dem die Instanzmethode aufgerufen wird. Anschließend werden alle Parameter in aufeinanderfolgenden lokalen Variablen übergeben, beginnend mit der lokalen Variable 1.

Beispiel: *Default Constructor* In Java Bytecode

Ein *Constructor* welcher keine expliziten Parameter hat und nur den super Konstruktor aufruft.

```
// Method descriptor ()V
// Stack: 1, Locals: 1
public Main();
  0  aload_0 [this]
  1  invokespecial java.lang.Object()
  4  return
```

Die Zeilennummern und die Informationen über die lokalen Variablen ist optional und wird nur für Debugging Zwecke benötigt.

```
Line numbers:          [pc: 0, line: 9]
Local variable table: [pc: 0, pc: 5]  local: this
                                   index: 0
                                   type:  de.dhbw.simplesecurepp.Main
```

Es gibt weitere Metainformationen, die „nur“ für Debugging-Zwecke benötigt werden, z.B. Informationen über die ursprünglich Quelle des Codes oder die sogenannte "Local Variable Type Table" in Hinblick auf generische Typinformationen. Solche Informationen werden häufig vor Auslieferung entfernt bzw. nicht hineinkompiliert.

Beispiel: Aufruf einer komplexeren Methode

```
// Method descriptor ([Ljava/lang/String;)V
// Stack: 5, Locals: 8
public static void main(java.lang.String[] args) throws ...;
  0  aload_0 [args]
  1  arraylength
  2  iconst_2
  3  if_icmpeq 74 // integer comparison for equality
  6  getstatic java.lang.System.err : java.io.PrintStream
  9  ldc <String "SimpleSecure++">
 11  invokevirtual java.io.PrintStream.println(java.lang.String) : void
  ...
```

4. VERSCHLÜSSELUNG VON DATEN

Prof. Dr. Michael Eichberg

Alternativen zur Speicherung von Passwörtern

In einigen Anwendungsgebieten ist es möglich auf das explizite Speichern von Passwörtern ganz zu verzichten[*].

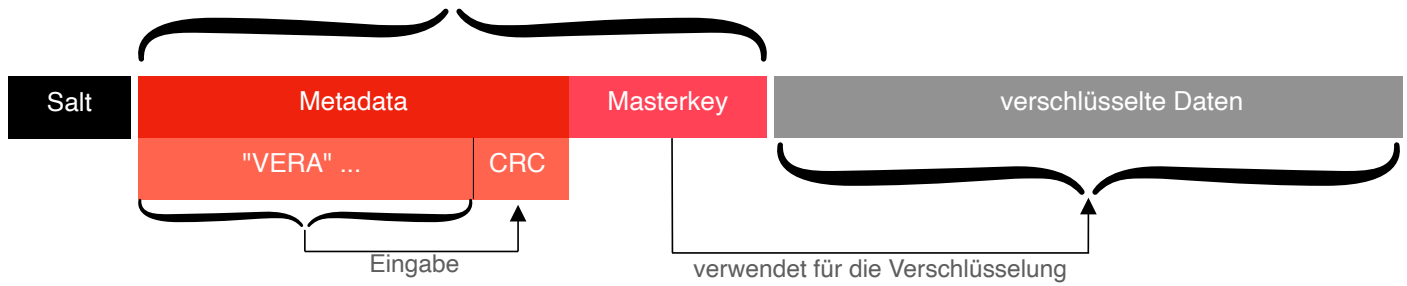
Stattdessen wird z.B. einfach versucht das Ziel zu entschlüsseln und danach evaluiert ob das Passwort (vermutlich) das Richtige war.

Kann darauf verzichtet werden zu überprüfen ob das Passwort korrekt war, dann sind keine Metainformationen notwendig und die verschlüsselte Datei kann genau so groß sein wie die unverschlüsselte Datei.

[*] Bei einer Verschlüsselung mit OpenSSL wird das Passwort nicht gespeichert.

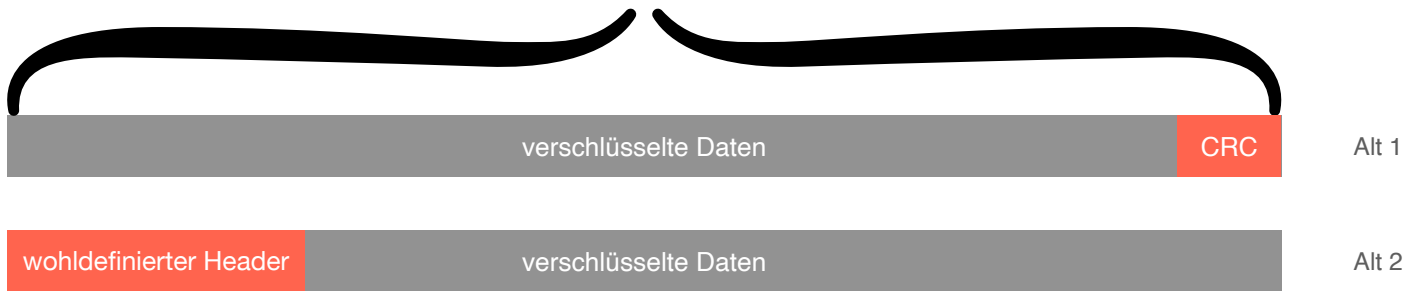
Schematische Darstellung der Verschlüsselung von Containern (z. B. Veracrypt)

Der Schlüssel wird mit Hilfe bekannter Schlüsselableitungsfunktionen aus dem Nutzerpasswort berechnet.



Generische Dateiverschlüsselung ohne explizite Speicherung des Passworts

verschlüsselt mit einem vom Passwort des Nutzers abgeleiteten Schlüssel



Die Prüfsumme (CRC) oder der wohldefinierte Header werden zur Validieren des Passworts verwendet.

Wichtig

Bleibe fokussiert!

Analysiere nur was notwendig ist.

Live Demo - Reversing SimpleSecure++

Gegeben

Programm: Secure++

Datei: Poem.enc

Exemplarische Verwendung zum Verschlüsseln

```
java -jar securepp-0.0.1.jar de.dhbw.securepp.Main \  
-p 'VielleichtIstEsRichtig-vielleichtAuchNICHT...' \  
-in Poem.txt -out Poem.enc
```

Aufgabe

Entschlüsseln Sie die Datei Poem.enc, die mit dem Program Secure++ verschlüsselt wurde.

Entschlüsseln Sie die Datei Poem.enc, die mit dem Program Secure++ verschlüsselt wurde.