Lecture 1: What is Secure Compilation?

Secure Compilation Seminar

Marco Patrignani

CISPA HELMHOLTZ-ZENTRUM i. G.

What is a Compiler

What is a Compiler

https://en.wikipedia.org/wiki/Compiler

What is a Compiler

https://en.wikipedia.org/wiki/Compiler

In this course:

- only care about the code generation phase

What is a Compiler

https://en.wikipedia.org/wiki/Compiler

In this course:

- only care about the code generation phase
- takes programs written in a source language S

What is a Compiler

https://en.wikipedia.org/wiki/Compiler

In this course:

- only care about the code generation phase
- takes programs written in a source language S
- output programs written in a target language T

What is a Compiler

https://en.wikipedia.org/wiki/Compiler

In this course:

- only care about the code generation phase
- takes programs written in a source language S
- output programs written in a **target language T**
- it is a function from S to **T**: $[\![\cdot]\!]^{\mathsf{S}}_{\mathbf{T}}$

Example #1: Insecure Compilation

```java
public class Account
  private int balance = 0;

  public void deposit( int amount )
    this.balance += amount;
```

Java source

Example #1: Insecure Compilation

```
public class Account
  private int balance = 0;

  public void deposit( int amount )
    this.balance += amount;
```

Java
source

No access to `balance` from outside `Account`

Example #1: Insecure Compilation

```java
public class Account
  private int balance = 0;

  public void deposit( int amount )
    this.balance += amount;
```

Java
source

No access to `balance` from outside `Account`

enforced by the language

Example #1: Insecure Compilation

```java
public class Account
  private int balance = 0;

  public void deposit( int amount )
    this.balance += amount;
```

Java
source

```c
typedef struct account_t {
  int balance = 0;
  void ( *deposit ) ( struct Account*, int ) =
      deposit_f;
} Account;

void deposit_f( Account* a, int amount ) {
  a->balance += amount;
  return;
}
```

C
target

Example #1: Insecure Compilation

```
publi
  priv

  publ
    th
```

Pointer arithmetic in C leads to security violation: undesired access to `balance`

Security is not preserved.

```
typedef struct account_t {
  int balance = 0;
  void ( *deposit ) ( struct Account*, int ) =
      deposit_f;
} Account;

void deposit_f( Account* a, int amount ) {
  a->balance += amount;
  return;
}
```

Secure Compilation

- **Q:** what does it mean to preserve security properties across compilation?

Secure Compilation

- **Q:** what does it mean to preserve security properties across compilation?
- long standing question

Secure Compilation

- **Q:** what does it mean to preserve security properties across compilation?

- long standing question
- many anwers have been given, we focus on the formal ones

Secure Compilation

- **Q:** what does it mean to preserve security properties across compilation?

- long standing question
- many anwers have been given, we focus on the formal ones
- conceptually:

  "take what was secure in the source and preserve it in the target"

Secure Compilation

- **Q:** what does it mean to preserve security properties across compilation?

Even more questions!

- how do we identify (or specify) what is secure in the source?
- how do we preserve the meaning of a security property?

Secure Compilation

- **Q:** what does it mean to preserve security properties across compilation?

Even more questions!

- how do we identify (or specify) what is secure in the source?
- how do we preserve the meaning of a security property?

anwers provided in this seminar

Example #2: Confidentiality

Confidential: adjective

spoken, written, acted on, etc., in strict privacy or secrecy; secret:

Example #2: Confidentiality

Confidential: adjective

> *spoken, written, acted on, etc., in strict privacy or secrecy; secret:*

```
private secret : Int = 0;

public setSecret( ) : Int {
  secret = 0;
  return 0;
}
```

Java

source

Example #2: Confidentiality

Confidential: adjective

spoken, written, acted on, etc., in strict privacy or secrecy; secret:

```
private secret : Int = 0;

public setSecret( ) : Int {
  secret = 0;
  return 0;
}
```

Java source

- **Q:** how do we know that `secret` is confidential?

Program Equivalence

- a possible answer to the questions before

Program Equivalence

- a possible answer to the questions before
- and to many more questions posed about programming languages

Quiz #1: Are these Equivalent Programs?

```
public Bool getTrue( x : Bool )
  return true;
```

```
public Bool getTrue( x : Bool )
  return x or true;
```

```
public Bool getTrue( x : Bool )
  return x and false;
```

```
public Bool getTrue( x : Bool )
  return false;
```

```
public Bool getFalse( x : Bool )
  return x and true;
```

- P1
- P2
- P3
- P4
- P5

Quiz #1: Are these Equivalent Programs?

```
public Bool getTrue( x : Bool )
  return true;
```

```
public Bool getTrue( x : Bool )
  return x or true;
```

```
public Bool getTrue( x : Bool )
  return x and false;
```

```
public Bool getTrue( x : Bool )
  return false;
```

```
public Bool getFalse( x : Bool )
  return x and true;
```

- P1
- P2

$\simeq_{ctx}$

- P3
- P4

$\simeq_{ctx}$

- P5

Quiz #1: Are these Equivalent Programs?

```
public Bool getTrue( x : Bool )
  return
```

```
public
  return
```

```
public
  return
```

```
public Bool getTrue( x : Bool )
  return false;
```

```
public Bool getFalse( x : Bool )
  return x and true;
```

P1

P4

P5

Program equivalences (generally) are:

- reflexive
- transitive
- symmetric

aka: relations

Program Equivalence

- **Q:** When are two programs equivalent?

Program Equivalence

- **Q:** When are two programs equivalent?
- When they behave the same

Program Equivalence

- **Q:** When are two programs equivalent?
- When they behave the same even if they are different

Program Equivalence

- **Q:** When are two programs equivalent?

- When they behave the same even if they are different
- Semantics (behaviour) VS Syntax (outlook)

Program Equivalence

- **Q:** When are two programs equivalent?

- When they behave the same even if they are different
- Semantics (behaviour) VS Syntax (outlook)
- we care about the former, not the latter!

Program Equivalence

Defining a security property using program equivalence:
to find two programs that, albeit syntactically different, both behave in a way that respects the property, no matter how they are used.

Example #4: Confidentiality as P.Eq.

Example #4: Confidentiality as P.Eq.

```
private secret : Int = 0;

public setSecret( ) : Int {
  secret = 0;
  return 0;
}
```

```
private secret : Int = 0;

public setSecret( ) : Int {
  secret = 1;
  return 0;
}
```

Example #4: Confidentiality as P.Eq.

```
private secret : Int = 0;

publi
  secr
  retu
}
```

```
priva

publi
  secr
  retu
}
```

With a Java-like semantics, `secret` is never accessed from outside.
With a C-like semantics, `secret` can be accessed from outisde.

Example #4: Confidentiality as P.Eq.

```
private secret : Int = 0;

publi
  secr
  retu
}
```

```
priva

publi
  secr
  retu
}
```

With a Java-like semantics, `secret` is never accessed from outside.

With a C-like semantics, `secret` can be accessed from outisde.

The Language defines how to reason (it's what programmers already do!)

Example #5: Integrity as P.Eq.

```
public proxy( callback : Unit → Unit ) : Int {
  var secret = 0;
  callback();
  if ( secret == 0 ) {
    return 0;
  }
  retu
}
```

Integrity: internal consistency or lack of corruption in data.

```
publi
  var secret = 0;
  callback();
  return 0;
}
```

Example #5: Integrity as P.Eq.

```
public proxy( callback : Unit → Unit ) : Int {
  var secret = 0;
  callback();
  if ( secret == 0 ) {
    return 0;
  }
  retu
}
```

Integrity: internal consistency or lack of corruption in data.
Maintenance of invariants

```
publi
  var secret = 0;
  callback();
  return 0;
}
```

Example #5: Integrity as P.Eq.

```
public proxy( callback : Unit → Unit ) : Int {
  var secret = 0;
  callback();
  if ( secret == 0 ) {
    return 0;
  }
  return 1;
}
```

```
public proxy( callback : Unit → Unit ) : Int {
  var secret = 0;
  callback();
  return 0;
}
```

Example #6: Memory Allocation as P.Eq.

```
public newObjects( ) : Object {
  var x = new Object();
  var y = new Object();
  return x;
}
```

```
public newObjects( ) : Object {
  var x = new Object();
  var y = new Object();
  return y;
}
```

Example #6: Memory Allocation as P.Eq.

```
public newObjects( ) : Object {
  var x = new Object();
  var y = new Object();
  retu
}
```

```
publi
  var
  var
  return y;
}
```

Guessing addresses in memory leads to common exploits: ROP, return to libc, violation of ASLR …

Expressing Program Equivalence

This ($\simeq_{ctx}$) is called

Contextual Equivalence

Expressing Program Equivalence

This ($\simeq_{ctx}$) is called

Contextual Equivalence

(also, observational equivalence)

Contextual Equivalence (CEQ)

Two programs are equivalent if no matter what external observer interacts with them that observer cannot distinguish the programs.

Contextual Equivalence (CEQ)

Two programs are equivalent if no matter what external observer interacts with them that observer cannot distinguish the programs.

$$P_1 \simeq_{ctx} P_2 \stackrel{\text{def}}{=} \forall P.\ P; P_1 {\downarrow} P; P_2$$

Contextual Equivalence (CEQ)

Two programs are equivalent if no matter what external observer interacts with them that observer cannot distinguish the programs.

$$P_1 \simeq_{ctx} P_2 \stackrel{\text{def}}{=} \forall P.\ P; P_1 \downarrow P; P_2$$

Contextual Equivalence (CEQ)

Two programs are equivalent if no matter what external observer interacts with them that observer cannot distinguish the programs.

$$P_1 \simeq_{ctx} P_2 \stackrel{\text{def}}{=} \forall\, P.\; P; P_1 {\downarrow} P; P_2$$

Contextual Equivalence (CEQ)

Two programs are equivalent if no matter what external observer interacts with them that observer cannot distinguish the programs.

$$P_1 \simeq_{ctx} P_2 \stackrel{\text{def}}{=} \forall P.\ P; P_1 {\downarrow} P; P_2$$

Contextual Equivalence (CEQ)

Two programs are equivalent if no matter what external observer interacts with them that observer cannot distinguish the programs.

$$P_1 \simeq_{ctx} P_2 \stackrel{\text{def}}{=} \forall P.\ P; P_1 {\downarrow} P; P_2$$

- the external observer is generally called context
- it is a program, written in the same language as P_1 and P_2
- it is the same program P interacting with both P_1 and P_2 in two different runs
- so it cannot express out of language attacks (e.g., side channels)
- often we write it $\mathfrak{C}$ or $\mathbb{C}$

Conte

- the external observer is generally called context
- it is a program, written in the same language as P_1 and P_2

Two ... at
exter...
obse...

- interaction means link and run together (like a library)
- often we write it $\mathfrak{C}[P_1]$

- so it cannot express out of language attacks (e.g., side channels)
- often we write it $\mathfrak{C}$ or $\mathbb{C}$

- distinguishing means: terminate with different values
- the observer basically asks the question: *is this program* P_1*?*
- if the observer can find a way to distinguish P_1 from P_2, it will return `true`, otherwise `false`

- distinguishing means: terminate with different values
- the observer basically asks the question: *is this program* P_1*?*
- if the observer can find a way to distinguish P_1 from P_2, it will return `true`, otherwise `false`
- often we use *divergence* and *termination* as opposed to this boolean termination

Example #7: CEQ

```
private secret : Int = 0; //P1
public setSecret( ) : Int {
  secret = 0;
  return 0;
}
```

```
private secret : Int = 0; //P2
public setSecret( ) : Int {
  secret = 1;
  return 0;
}
```

Example #7: CEQ

```
private secret : Int = 0; //P1
public setSecret( ) : Int {
  secret = 0;
  return 0;
}
```

Java

```
private secret : Int = 0; //P2
public setSecret( ) : Int {
  secret = 1;
  return 0;
}
```

Java

```
// Observer P in Java
public static isItP1( ) : Bool {
  Secret.getSecret();
  ...
}
```

Java

Example #8: CEQ

C

```
typedef struct secret { // P1
  int secret = 0;
  void ( *setSec ) ( struct Secret* ) = setSec;
} Secret;
void setSec( Secret* s ) { s->secret = 0; return; }
```

C

```
typedef struct secret { // P2
  int secret = 0;
  void ( *setSec ) ( struct Secret* ) = setSec;
} Secret;
void setSec( Secret* s ) { s->secret = 1; return; }
```

C

```
// Observer P in C
int isItP1( ){
  struct Secret x;
  sec = &x + sizeof(int);
  if *sec == 0 then return true else return false
}
```

Inequivalences as Security Violations

- if the programs are not equivalent ($\not\simeq_{ctx}$) then the intended security property is violated

Inequivalences as Security Violations

- if the programs are not equivalent ($\not\simeq_{ctx}$) then the intended security property is violated
- this hardly happens as *source* languages are high level

Inequivalences as Security Violations

When does inequivalences escape the programmer's reasoning?

Inequivalences as Security Violations

When does inequivalences escape the programmer's reasoning?

1. if languages have complex features

Inequivalences as Security Violations

When does inequivalences escape the programmer's reasoning?

1. if languages have complex features
2. if there are more languages involved (e.g., multiple target languages)

Preserving Equivalences in Compilation

Back to our question ...

- **Q:** what does it mean to preserve security properties across compilation?

Preserving Equivalences in Compilation

Back to our question . . .

- **Q:** what does it mean to preserve security properties across compilation?

A possible answer:

- Given source equivalent programs (which have a security property), compile them into equivalent target programs

Preserving Equivalences in Compilation

- Assumption 1: the security property is captured in the source by program equivalence

Preserving Equivalences in Compilation

- Assumption 1: the security property is captured in the source by program equivalence
- **Crucial**: being equivalent in the target means contextual equivalence w.r.t. target observers (i.e., target programs)

Preserving Equivalences in Compilation

- Assumption 1: the security property is captured in the source by program equivalence
- **Crucial**: being equivalent in the target means contextual equivalence w.r.t. target observers (i.e., target programs)
- These are the attackers in the secure compilation setting

Fully Abstract Compilation

Fully Abstract Compilation

A compiler is secure if, given source equivalent programs, it compiles them into equivalent target programs

$$\llbracket \cdot \rrbracket^{\mathsf{S}}_{\mathbf{T}} \text{ is FAC\#1} \stackrel{\text{def}}{=} \forall \mathsf{P}_1, \mathsf{P}_2$$

$$\text{if } \mathsf{P}_1 \simeq_{ctx} \mathsf{P}_2$$

$$\text{then } \llbracket \mathsf{P}_1 \rrbracket^{\mathsf{S}}_{\mathbf{T}} \simeq_{ctx} \llbracket \mathsf{P}_2 \rrbracket^{\mathsf{S}}_{\mathbf{T}}$$

Fully Abstract Compilation

A compiler is secure if, given source equivalent programs, it compiles them into equivalent target programs

$$\llbracket \cdot \rrbracket^{\mathsf{S}}_{\mathbf{T}} \text{ is FAC\#1} \stackrel{\text{def}}{=} \forall \mathsf{P}_1, \mathsf{P}_2$$

$$\text{if } \mathsf{P}_1 \simeq_{ctx} \mathsf{P}_2$$

$$\text{then } \llbracket \mathsf{P}_1 \rrbracket^{\mathsf{S}}_{\mathbf{T}} \simeq_{ctx} \llbracket \mathsf{P}_2 \rrbracket^{\mathsf{S}}_{\mathbf{T}}$$

Fully Abstract Compilation

A compiler is secure if, given source equivalent programs, it compiles them into equivalent target programs

$$\llbracket \cdot \rrbracket^{\mathsf{S}}_{\mathbf{T}} \text{ is FAC\#1} \stackrel{\text{def}}{=} \forall \mathsf{P}_1, \mathsf{P}_2$$

$$\text{if } \mathsf{P}_1 \simeq_{ctx} \mathsf{P}_2$$

$$\text{then } \llbracket \mathsf{P}_1 \rrbracket^{\mathsf{S}}_{\mathbf{T}} \simeq_{ctx} \llbracket \mathsf{P}_2 \rrbracket^{\mathsf{S}}_{\mathbf{T}}$$

Fully Abstract Compilation

A compiler is secure if, given source equivalent programs, it compiles them into equivalent target programs

$$[\![\cdot]\!]^{\mathsf{S}}_{\mathbf{T}} \text{ is FAC\#1} \overset{\text{def}}{=} \forall \mathsf{P}_1, \mathsf{P}_2$$

$$\text{if } \mathsf{P}_1 \simeq_{ctx} \mathsf{P}_2$$

$$\text{then } [\![\mathsf{P}_1]\!]^{\mathsf{S}}_{\mathbf{T}} \simeq_{ctx} [\![\mathsf{P}_2]\!]^{\mathsf{S}}_{\mathbf{T}}$$

Fully Abstract Compilation

A compiler is secure if, given source equivalent programs, it compiles them into equivalent target programs

$$
\begin{aligned}
[\![\cdot]\!]^{\mathsf{S}}_{\mathbf{T}} \text{ is FAC\#1} &\stackrel{\text{def}}{=} \forall \mathsf{P}_1, \mathsf{P}_2 \\
&\text{if } \mathsf{P}_1 \simeq_{ctx} \mathsf{P}_2 \\
&\text{then } [\![\mathsf{P}_1]\!]^{\mathsf{S}}_{\mathbf{T}} \simeq_{ctx} [\![\mathsf{P}_2]\!]^{\mathsf{S}}_{\mathbf{T}}
\end{aligned}
$$

Fully Abstract Compilation

A compiler is secure if, given source equivalent programs, it compiles them into equivalent target programs

$$\llbracket\cdot\rrbracket^{\mathsf{S}}_{\mathbf{T}} \text{ is FAC\#1} \stackrel{\text{def}}{=} \forall \mathsf{P}_1, \mathsf{P}_2$$

$$\text{if } \mathsf{P}_1 \simeq_{ctx} \mathsf{P}_2$$

$$\text{then } \llbracket\mathsf{P}_1\rrbracket^{\mathsf{S}}_{\mathbf{T}} \simeq_{ctx} \llbracket\mathsf{P}_2\rrbracket^{\mathsf{S}}_{\mathbf{T}}$$

Right?

Fully Abstract Compilation

Wrong.

[1] Abadi ’99

Fully Abstract Compilation

Wrong.

An empty translation would fit FAC#1!

[1]Abadi '99

Fully Abstract Compilation

Wrong.

An empty translation would fit FAC#1!

We need the compiler also to be correct.

Roughly, turn $\Rightarrow$ into a $\iff$:[1]

$$[\![\cdot]\!]^{\mathsf{S}}_{\mathbf{T}} \text{ is FAC} \stackrel{\text{def}}{=} \forall \mathsf{P}_1, \mathsf{P}_2$$

$$\mathsf{P}_1 \simeq_{ctx} \mathsf{P}_2 \iff [\![\mathsf{P}_1]\!]^{\mathsf{S}}_{\mathbf{T}} \simeq_{ctx} [\![\mathsf{P}_2]\!]^{\mathsf{S}}_{\mathbf{T}}$$

[1]Abadi ’99

Fully Abstract Compilation

Wrong.

An empty translation would fit FAC#1!

We need the compiler also to be correct.

Roughly, turn $\Rightarrow$ into a $\iff$:[1]

$$[\![\cdot]\!]^{\mathsf{S}}_{\mathbf{T}} \text{ is FAC} \stackrel{\text{def}}{=} \forall \mathsf{P}_1, \mathsf{P}_2$$

$$\mathsf{P}_1 \simeq_{ctx} \mathsf{P}_2 \iff [\![\mathsf{P}_1]\!]^{\mathsf{S}}_{\mathbf{T}} \simeq_{ctx} [\![\mathsf{P}_2]\!]^{\mathsf{S}}_{\mathbf{T}}$$

Note: $\Leftarrow$ does not mean compiler correctness in the general sense, but it's a consequence

[1] Abadi '99

Remarks on Fully Abstract Compilation

- widely adopted since 1999

Remarks on Fully Abstract Compilation

- widely adopted since 1999
- only preserves security property expressed as program equivalence

Remarks on Fully Abstract Compilation

- widely adopted since 1999
- only preserves security property expressed as program equivalence
- not the silver bullet: the papers will present shortcomings of fully abstract compilation

Conclusion

- program equivalences can be used to define security properties
- preserving (and reflecting) equivalences can be used to define a secure compiler

Further Reading

Short(ish) and high-level(ish):

- Martin Abadi. 1999. Protection in programming-language translations. In Secure Internet programming
- Andrew Kennedy. 2006. Securing the .NET Programming Model. Theoretical Computer Science 364 (2006)
- Joachim Parrow. 2014. General conditions for Full Abstraction. Math Struct Comp Science (2014)
- Daniele Gorla and Uwe Nestman. 2014. Full Abstraction for Expressiveness: History, Myths and Facts. Math Struct Comp Science (2014)