Hadrien Barral
Georges-Axel Jaloyan

École Normale Supérieure,
PSL Research University

3MOJI SHELLCODING:
🛠️️, 🧙‍♂️️, AND 😈️️
WHO WE ARE

- **Hadrien Barral**: Hacker 🕵️‍♂️. I like hacking stuff.
- **Georges-Axel Jaloyan**: PhD student. I reverse open source binaries.
- This is our third/second talk at DEFCON.
WHAT THIS IS ABOUT

- Some generic 🛠️ and methods
- ... with lots of emoji 🤖
- ... to write 😳 shellcodes.
1. WHY WOULD ANYONE WRITE CONSTRAINED SHELLCODES?
SHELLCODE 101
For those hiding in the back

- Code that you wrote (or found) in the target’s memory
- That gives you some power (e.g. pop a shell)
- That you can jump to using some vuln (e.g. buffer overflow, UAF...)

Usual scenario: send a carefully crafted string to the target and profit.
SHELLCODE 102
Not as convenient as you may think

- If treated as a C string, no \x00
- If treated as input, no whitespace
- Other constraints (regex)
- Breaks easily (string escaping)
- Does not really look legit 😞
Constrained shellcoding on x86 is easy:

<table>
<thead>
<tr>
<th>ASCII</th>
<th>IA32</th>
</tr>
</thead>
<tbody>
<tr>
<td>'P'</td>
<td>push %eax</td>
</tr>
<tr>
<td>'h'</td>
<td>push imm32</td>
</tr>
<tr>
<td>'X'</td>
<td>pop %eax</td>
</tr>
<tr>
<td>'a'</td>
<td>popa</td>
</tr>
<tr>
<td>'A'</td>
<td>inc %eax</td>
</tr>
<tr>
<td>'H'</td>
<td>dec %eax</td>
</tr>
<tr>
<td>'t'</td>
<td>je short disp8</td>
</tr>
<tr>
<td>'u'</td>
<td>jnz short disp8</td>
</tr>
<tr>
<td>'v'</td>
<td>jbe short disp8</td>
</tr>
<tr>
<td>'8'</td>
<td>cmp r/m8 r8</td>
</tr>
<tr>
<td>'0'</td>
<td>xor r/m8 r8</td>
</tr>
</tbody>
</table>
Constrained shellcoding on x86 is easy:

- push/pop/inc/dec have single letter instructions!

<table>
<thead>
<tr>
<th>ASCII</th>
<th>IA32</th>
</tr>
</thead>
<tbody>
<tr>
<td>'P'</td>
<td>push %eax</td>
</tr>
<tr>
<td>'h'</td>
<td>push imm32</td>
</tr>
<tr>
<td>'X'</td>
<td>pop %eax</td>
</tr>
<tr>
<td>'a'</td>
<td>popa</td>
</tr>
<tr>
<td>'A'</td>
<td>inc %eax</td>
</tr>
<tr>
<td>'H'</td>
<td>dec %eax</td>
</tr>
<tr>
<td>'t'</td>
<td>je short disp8</td>
</tr>
<tr>
<td>'u'</td>
<td>jnz short disp8</td>
</tr>
<tr>
<td>'v'</td>
<td>jbe short disp8</td>
</tr>
<tr>
<td>'8'</td>
<td>cmp r/m8 r8</td>
</tr>
<tr>
<td>'0'</td>
<td>xor r/m8 r8</td>
</tr>
</tbody>
</table>
Constrained shellcoding on x86 is easy:

- push/pop/inc/dec have single letter instructions!
- jmp and cmp are available

<table>
<thead>
<tr>
<th>ASCII</th>
<th>IA32</th>
</tr>
</thead>
<tbody>
<tr>
<td>'P'</td>
<td>push %eax</td>
</tr>
<tr>
<td>'h'</td>
<td>push imm32</td>
</tr>
<tr>
<td>'X'</td>
<td>pop %eax</td>
</tr>
<tr>
<td>'a'</td>
<td>popa</td>
</tr>
<tr>
<td>'A'</td>
<td>inc %eax</td>
</tr>
<tr>
<td>'H'</td>
<td>dec %eax</td>
</tr>
<tr>
<td>'t'</td>
<td>je short disp8</td>
</tr>
<tr>
<td>'u'</td>
<td>jnz short disp8</td>
</tr>
<tr>
<td>'v'</td>
<td>jbe short disp8</td>
</tr>
<tr>
<td>'8'</td>
<td>cmp r/m8 r8</td>
</tr>
<tr>
<td>'0'</td>
<td>xor r/m8 r8</td>
</tr>
</tbody>
</table>
Constrained shellcoding on x86 is easy:

- **push/pop/inc/dec** have single letter instructions!
- **jmp** and **cmp** are available
- **xor** with many operands are available

<table>
<thead>
<tr>
<th>ASCII</th>
<th>IA32</th>
</tr>
</thead>
<tbody>
<tr>
<td>'P'</td>
<td>push %eax</td>
</tr>
<tr>
<td>'h'</td>
<td>push imm32</td>
</tr>
<tr>
<td>'X'</td>
<td>pop %eax</td>
</tr>
<tr>
<td>'a'</td>
<td>popa</td>
</tr>
<tr>
<td>'A'</td>
<td>inc %eax</td>
</tr>
<tr>
<td>'H'</td>
<td>dec %eax</td>
</tr>
<tr>
<td>'t'</td>
<td>je short disp8</td>
</tr>
<tr>
<td>'u'</td>
<td>jnz short disp8</td>
</tr>
<tr>
<td>'v'</td>
<td>jbe short disp8</td>
</tr>
<tr>
<td>'8'</td>
<td>cmp r/m8 r8</td>
</tr>
<tr>
<td>'0'</td>
<td>xor r/m8 r8</td>
</tr>
</tbody>
</table>
Constrained shellcoding on x86 is easy:

- `push/pop/inc/dec` have single letter instructions!
- `jmp` and `cmp` are available
- `xor` with many operands are available
- 🤖 for x86-64: just prefix `0x48 (= H)` to every instruction

<table>
<thead>
<tr>
<th>ASCII</th>
<th>IA32</th>
</tr>
</thead>
<tbody>
<tr>
<td>'P'</td>
<td><code>push %eax</code></td>
</tr>
<tr>
<td>'h'</td>
<td><code>push imm32</code></td>
</tr>
<tr>
<td>'X'</td>
<td><code>pop %eax</code></td>
</tr>
<tr>
<td>'a'</td>
<td><code>popa</code></td>
</tr>
<tr>
<td>'A'</td>
<td><code>inc %eax</code></td>
</tr>
<tr>
<td>'H'</td>
<td><code>dec %eax</code></td>
</tr>
<tr>
<td>'t'</td>
<td><code>je short disp8</code></td>
</tr>
<tr>
<td>'u'</td>
<td><code>jnz short disp8</code></td>
</tr>
<tr>
<td>'v'</td>
<td><code>jbe short disp8</code></td>
</tr>
<tr>
<td>'8'</td>
<td><code>cmp r/m8 r8</code></td>
</tr>
<tr>
<td>'0'</td>
<td><code>xor r/m8 r8</code></td>
</tr>
</tbody>
</table>
CRAZY PEOPLE EVEN MADE ENGLISH SHELLCODES

Mason et al., CCS, 2009
ALPHANUMERIC SHELLCODING: TAKING RISCs

RISC – *reduced* instruction set computer...

- No single character instructions anymore
- Few addressing modes – in particular no memory to memory
- Heavy constraints on operands

Previous technique does not work on RISC architectures! 😱
RISC – reduced instruction set computer...

- No single character instructions anymore
- Few addressing modes – in particular no memory to memory
- Heavy constraints on operands

Previous technique does not work on RISC architectures! 😱

Compilation / Emulation / Unpacking
THE "COMPILATION" WAY

Idea:
- Compile assembly code to a constrained instruction set

Pros:
- Feasible for one-instruction set computers (e.g. Movfuscator on x86)

Cons:
- Does not work when the constraints are mainly on the operands and not on the opcodes
- Nobody wants to devote their life to writing such a compiler
THE "COMPILATION" WAY

Idea:
- Compile assembly code to a constrained instruction set

Pros:
- Feasible for one-instruction set computers (e.g. Movfuscator on x86)

Cons:
- Does not work when the constraints are mainly on the operands and not on the opcodes
- ...nobody wants to devote their life to writing such a compiler 😞
THE "EMULATION" WAY
Younan et al., Phrack 66, 2009

Idea:
- Write an interpreter for some language

Pros:
- Reusable for different payloads
- e.g. Younan's ARMv7 alphanumeric Brainfuck interpreter

Cons:
Idea:
- Write an interpreter for some language

Pros:
- Reusable for different payloads
- e.g. Younan’s ARMv7 alphanumeric Brainfuck interpreter

Cons:
- Harmfulness now relies on the interpreter...
The "Unpacking" Way
Barral et al., ISPEC 2016

Idea:
- Encode payload in a constraint-compliant way (e.g. alphanumerically)
THE "UNPACKING" WAY
Barral et al., ISPEC 2016

Idea:
- Encode payload in a constraint-compliant way (e.g. alphanumerically)
- Identify high-level constraint-compliant constructs (zeroing/increasing registers, ...)
THE "UNPACKING" WAY
Barral et al., ISPEC 2016

Idea:

- Encode payload in a constraint-compliant way (e.g. alphanumerically)
- Identify high-level constraint-compliant constructs (zeroing/increasing registers, ...)
- Use this to write a minimal unpacker that decodes and executes payload 🙀
ALPHANUMERIC IS "SOLVED"

Even works for RISC-V!

The ABC of Next-Gen Shellcoding, DEF CON 27

(protip, it is our previous talk)
ONCE UPON A TIME... (TRUE STORY)

Can you emoji shellcode?

Nope, because XYZ!

@BloodyTangerine

Shit XYZ is wrong!

Hadrien

I now haz to write emoji shellcode

WTF is this?

With all those bugs, how does it even work?

Georges-Axel

Let's clean this shit together

Georges-Axel

Hadrien
2. INSPECTOR GADGET
WHAT IS AN EMOJI?
WHAT IS AN EMOJI?

(or more likely a mess 😒 ...)

I 🖤 you

I 🤡 you

emoji Shellcoding: 🛠️, 🗝️, and 🤡 @ DEFCON

12 Aug 2022

17/42
WHAT IS AN EMOJI?

Unicode (1987) to the rescue

“Unicode is a standard for the consistent encoding, representation, and handling of text expressed in most of the world’s writing systems.”

— Wikipedia
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A,
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま,
WHAT IS AN EMOJI?
RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, 🇸,

Definition:
If Unicode says it is a qualified Emoji, then it is an Emoji!

In UTF-8, emojis at least 3 bytes, at most 35.
And they add new emojis every year! Currently at Unicode version 14.
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, ♢, ♧,
WHAT IS AN EMOJI?
RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, ♠, 🎁, 😊,
WHAT IS AN EMOJI?
RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, パ, ♠️, 😊, 🁒, 🁜,
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, ℹ, 🎠, 😄, 🕛, 🟠, 🌏,

Definition: If Unicode says it is a qualified Emoji, then it is an Emoji! In UTF-8, emojis at least 3 bytes, at most 35. And they add new emojis every year! Currently at Unicode version 14.
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, 岩, ♠, 😄, 🌐, 🕊️, 🗠️, 🪵,
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, alsy, 😊, 🔐, 🌐, ❌,

Definition: If Unicode says it is a qualified Emoji, then it is an Emoji!

In UTF-8, emojis at least 3 bytes, at most 35.

And they add new emojis every year! Currently at Unicode version 14.
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, あ, 🎉, 😊, 🏇, 🌐, 🍃, 👑,
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, 🍀, 😄, 🟥, 🌃, 🌁, 🌡️, 🛡️,
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, 仓, 🕶️, 😄, 🌀🌄, 🛍️, 🤚

Definition:
If Unicode says it is a qualified Emoji, then it is an Emoji!
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま, あ, け, 😊, 🌞, 🌙,

Definition:

If Unicode says it is a qualified Emoji, then it is an Emoji!

In UTF-8, emojis at least 3 bytes 😊,
WHAT IS AN EMOJI?

RTFM: https://unicode.org/reports/tr51/

In Unicode, you’ll find:

A, ま,  ipv, ♦, 😊, Gesture, 𒈙

Definition:

If Unicode says it is a qualified Emoji, then it is an Emoji!

In UTF-8, emojis at least 3 bytes 😊, at most 35 😘.
WHAT IS AN EMOJI?

**RTFM:** [https://unicode.org/reports/tr51/](https://unicode.org/reports/tr51/)

In Unicode, you’ll find:

A, ま, ippers, 😊, 🂶, 🌐, 🌺, 🎇, 🧜‍♀️

**Definition:**

If Unicode says it is a *qualified Emoji*, then it is an Emoji!

In UTF-8, emojis at least 3 bytes 😊, at most 35 💘.

And they add new emojis every year! Currently at Unicode version 14.
EXCUSE ME, DO YOU HAVE 2 MINUTES TO TALK ABOUT

RISC-V®
EXCUSE ME, DO YOU HAVE 2 MINUTES TO TALK ABOUT

Features:
- Architecture of the future
- Simple, clean RISC Instruction Set Architecture
- Open Source ISA & Open Hardware*
- 2 and 4 byte instructions*, little-endian

*: Conditions may apply
EXCUSE ME, DO YOU HAVE 2 MINUTES TO TALK ABOUT

**RISC-V**

Features:
- Architecture of the future present
- Simple, clean RISC Instruction Set Architecture
- Open Source ISA & Open Hardware*
- 2 and 4 byte instructions*, little-endian

*:Conditions may apply
EXCUSE ME, DO YOU HAVE 2 MINUTES TO TALK ABOUT

RISC-V

Features:
- Architecture of the future present
- Simple, clean RISC Instruction Set Architecture
- Open Source ISA & Open Hardware*
- 2 and 4 byte instructions*, little-endian

*: Conditions may apply
HOW CAN I EXECUTE MY EMOJI?
HOW CAN I EXECUTE MY EMOJI?

- **Alphanumeric x86**: ‘A’: `inc %eax`
  
  `0x41`

- **Alphanumeric ARM&RISCV**: `4A0s`
  
  `csrc mip,sp
  0x34413073`

- **Emoji RISCV**: `0xE29D93`

We are stuck!
HOW CAN I EXECUTE MY EMOJI?

- ✔ Alphanumeric x86: ‘A’: \texttt{inc \%eax}
  
  0x41

- ✔ Alphanumeric ARM&RISCV: ‘4A0s’: \texttt{csrc mip,sp}
  
  0x34413073
HOW CAN I EXECUTE MY EMOJI?

- ✔ Alphanumeric x86: ‘A’: inc %eax
  0x41

- ✔ Alphanumeric ARM&RISCV: ‘4A0s’: csrd mip,sp
  0x34413073

- ❔ Emoji RISCV:
HOW CAN I EXECUTE MY EMOJI?

✅ Alphanumeric x86: ‘A’: `inc %eax`  
0x41

✅ Alphanumeric ARM&RISCV: ‘4A0s’: `csrc mip,sp`  
0x34413073

❌ Emoji RISCV: ‘?’: only 10 instructions  
0xE29D93
HOW CAN I EXECUTE MY EMOJI?

- ☑ Alphanumeric x86: ‘A’: `inc %eax`
  
  0x41

- ☑ Alphanumeric ARM&RISCV: ‘4A0s’: `csrc mip,sp`

  0x34413073

- ❌ Emoji RISCV: `??`: still very few instructions

  0xE29D93E29D93
HOW CAN I EXECUTE MY EMOJI?

- **✓** Alphanumeric x86: ‘A’: `inc %eax
0x41`

- **✓** Alphanumeric ARM&RISCV: ‘4A0s’: `csrc mip,sp
0x34413073`

- **✗** Emoji RISCV: ? ? ? ? ... : does not get better 😪
  `0xE29D93E29D93E29D93...`
HOW CAN I EXECUTE MY EMOJI?

- **✅ Alphanumeric x86:** ‘A’ : `inc %eax`
  
  0x41

- **✅ Alphanumeric ARM&RISCV:** ‘4A0s’ : `csrc mip,sp`
  
  0x34413073

- **🚫 Emoji RISCV:** ? ? ? ? ... : does not get better 😢
  
  0xE29D93E29D93E29D93...  

We are stuck!
auipc ra,0x979ff
THE INTUITION...

97 F0 9F 97

auipc ra,0x979ff
THE INTUITION...

97  F0  9F  97

auipc ra, 0x979ff
THE INTUITION...

F0 9F 86 97 F0 9F 97

auipc ra,0x979ff
THE INTUITION...

Jump

3moji Shellcoding:  🚀,  🔥,  and  🧪 @ DEFCON

12 Aug 2022  22/42
THE INTUITION...

Jump

<none>

F0 9F 86 97 F0 9F 97

auipc ra,0x979ff
Jump

<none>

\[ \text{auipc ra,0x979ff} \]

\[ \text{E2 9D 97} \]

\[ \text{F0 9F 97} \]
THE INTUITION...

Jump

<none>

add s11,s8

auipc ra,0x979ff
THE INTUITION...

Jump

<none>

F0 9F 86 E2 9D

add s11, s8

auipc ra, 0x979ff

F0 9F 97

93 EF B8 8F
THE INTUITION...
THE INTUITION...

Jump

add s11, s8

auipc ra, 0x979ff

ori t6, a7, -1797

Jump 3moji Hellcoding: 😂, 🌐, and 🤖 @ DEFCON

12 Aug 2022 22/42
THE INTUITION...

Jump

<none>

add s11,s8

auipc ra,0x979ff

ori t6,a7,-1797

bnez a5,+28

Jump
THE INTUITION...

Jump

<none>

add s11, s8

auipc ra, 0x979ff

ori t6, a7, -1797

bnez a5, +28
How do I generate all gadgets?
Return-Oriented Programming (Shacham et al., 2007):

1. Scan target binary for reusable code snippets (*gadgets*).
2. Build your shellcode with gadgets.
3. ???
4. Profit.
EMOJI SHELLCODING IS A CODE-REUSE PROBLEM
Source: Trust me, bro.

- Return-Oriented Programming (Shacham et al., 2007):
  1. Scan target binary for reusable code snippets (gadgets)
  2. Build your shellcode with gadgets
  3. ???
  4. Profit

- JIT spraying (Blazakis, 2010):
  create your own gadgets by controlling JITted bytecode.

```javascript
var y = (0x3c54d0d9 ^ 0x3c909058 ^ 0x3c59f46a ^
  ...
```

is turned into:
```
B8 D9D0543C MOV EAX,3C54D0D9
35 5890903C XOR EAX,3C909058
35 6AF4593C XOR EAX,3C59F46A
...```
EMOJI SHELLCODING IS A CODE-REUSE PROBLEM
Source: Trust me, bro.

Return-Oriented Programming (Shacham et al., 2007):
1. Scan target binary for reusable code snippets (*gadgets*)
2. Build your shellcode with gadgets
3. ???
4. Profit

JIT spraying (Blazakis, 2010):
create your own gadgets by controlling JITted bytecode.

Easy when you control most of the output...

```
var y = (0x3c54d0d9 ᵅ 0x3c909058 ᵅ 0x3c59f46a ᵅ ...)
```

is turned into:
```
B8 D9D0543C MOV EAX,3C54D0D9
35 5890903C XOR EAX,3C909058
35 6AF4593C XOR EAX,3C59F46A
...```
EMOJI SHELLCODING IS A CODE-REUSE PROBLEM

Great! How do I create gadgets from Emoji stream?
Emoji Shellcoding is a Code-Reuse Problem

Great! How do I create gadgets from Emoji stream?

Remember infinite monkey theorem?
EMOJI SHELLCODING IS A CODE-REUSE PROBLEM

Great! How do I create gadgets from Emoji stream?

Remember infinite monkey theorem?

They get emoji keyboards now!

⇒ Need to invent a new algorithm.
<table>
<thead>
<tr>
<th>Address</th>
<th>Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0</td>
<td>add s7, s7, s8</td>
</tr>
<tr>
<td>0x2</td>
<td></td>
</tr>
<tr>
<td>0x4</td>
<td></td>
</tr>
<tr>
<td>0x6</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
## EMOJI GADGET BUILDING (1ST CASE)

<table>
<thead>
<tr>
<th>0x0</th>
<th>0x2</th>
<th>0x4</th>
<th>0x6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>e2</td>
<td>9b</td>
<td>93</td>
<td>ef</td>
<td>b8</td>
</tr>
<tr>
<td>add s7, s7, s8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>add s7, s7, s8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bnez a5, +0x5e</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bnez a5, +0xe24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.byte 0xb8, 0x8f</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### EMOJI GADGET BUILDING (1ST CASE)

<table>
<thead>
<tr>
<th></th>
<th>e2</th>
<th>9b</th>
<th>93</th>
<th>ef</th>
<th>b8</th>
<th>8f</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>add s7,s7,s8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ori t6,a7,-1797</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0x0 0x2 0x4 0x6 ...
EMOJI GADGET BUILDING (1ST CASE)

<table>
<thead>
<tr>
<th>e2</th>
<th>9b</th>
<th>93</th>
<th>ef</th>
<th>b8</th>
<th>8f</th>
</tr>
</thead>
<tbody>
<tr>
<td>add s7,s7,s8</td>
<td>ori t6,a7,-1797</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 add s7,s7,s8; ori t6,a7,-1797
1 add s7, s7, s8; ori t6, a7, -1797
### EMOJI GADGET BUILDING (2ND CASE)

<table>
<thead>
<tr>
<th>e2</th>
<th>9b</th>
<th>b9</th>
<th>ef</th>
<th>b8</th>
<th>8f</th>
</tr>
</thead>
<tbody>
<tr>
<td>add s7,s7,s8</td>
<td>bnez a5,+0x5e</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 0x0  | 0x2  | 0x4  | 0x6  | ⋮   |

1. `add s7,s7,s8; ori t6,a7,-1797`
1 add s7,s7,s8; ori t6,a7,-1797
2 add s7,s7,s8; bnez a5,+0x5e; .byte 0xb8,0x8f
1. `add s7,s7,s8; ori t6,a7,-1797`
2. `add s7,s7,s8; bnez a5,+0x5e; .byte 0xb8,0x8f`
**EMOJI GADGET BUILDING (3RD CASE)**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e2</td>
<td>9b</td>
<td>84</td>
<td>c2</td>
</tr>
<tr>
<td>add s7, s7, s8</td>
<td>sw s1, 0(a3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0x0 0x2 0x4 0x6 ...

1. `add s7, s7, s8; ori t6, a7, -1797`
2. `add s7, s7, s8; bnez a5, +0x5e; .byte 0xb8, 0xf`
## EMOJI GADGET BUILDING (3RD CASE)

### Instructions

1. `add s7, s7, s8; ori t6, a7, -1797`
2. `add s7, s7, s8; bnez a5, +0x5e; .byte 0xb8, 0x8f`

### Table

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e2</td>
<td>9b</td>
<td>84</td>
<td>c2</td>
<td></td>
</tr>
<tr>
<td>add s7, s7, s8</td>
<td>sw s1, 0(a3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x0</td>
<td>0x2</td>
<td>0x4</td>
<td>0x6</td>
<td>⋱</td>
</tr>
</tbody>
</table>

---

3moji Shellcoding: 🎉, 🛠️, and 🤖 @ DEFCON

12 Aug 2022
# EMOJI GADGET BUILDING (3RD CASE)

<table>
<thead>
<tr>
<th>0x0</th>
<th>0x2</th>
<th>0x4</th>
<th>0x6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>add s7, s7, s8</td>
<td>sw s1, 0(a3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. `add s7, s7, s8; ori t6, a7, -1797`
2. `add s7, s7, s8; bnez a5, +0x5e; .byte 0xb8, 0x8f`
### Emoji Gadget Building (3rd Case)

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e2</td>
<td>9b</td>
<td>84</td>
<td>c2</td>
<td>a9</td>
<td>ef</td>
<td>b8</td>
<td>8f</td>
<td></td>
</tr>
<tr>
<td><strong>add s7,s7,s8</strong></td>
<td><strong>sw s1,0(a3)</strong></td>
<td><strong>bnez a5,+0xe24</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0x0 0x2 0x4 0x6 ... 0xe24

---

1. `add s7,s7,s8; ori t6,a7,-1797`
2. `add s7,s7,s8; bnez a5,+0x5e; .byte 0xb8,0x8f`
1. `add s7,s7,s8; ori t6,a7,-1797`
2. `add s7,s7,s8; bnez a5,+0xe5e; .byte 0xb8,0x8f`
3. `add s7,s7,s8; sw s1,0(a3); bnez a5,+0xe24; .byte 0xb8,0x8f`
## EMOJI GADGET BUILDING

<table>
<thead>
<tr>
<th>0x0</th>
<th>0x2</th>
<th>0x4</th>
<th>0x6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>e2</td>
<td>9b</td>
<td>84</td>
<td>c2</td>
<td>a9</td>
</tr>
<tr>
<td>add s7,s7,s8</td>
<td>sw s1,0(a3)</td>
<td>bnez a5,+0xe24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. `add s7,s7,s8; ori t6,a7,-1797`
2. `add s7,s7,s8; bnez a5,+0x5e; .byte 0xb8,0x8f`
3. `add s7,s7,s8; sw s1,0(a3); bnez a5,+0xe24; .byte 0xb8,0x8f`
4. (and so on...)
This method is totally generic

Algorithm follows derivations in the form \( S \rightarrow a \ T \)

This exactly describes a right-linear grammar!

“Any regular expression* can be converted to a right-linear grammar.”

*alphanumeric, email address, url, emojis, ...

— A.R. Patel, 1971

(adapting the tool is left to the reader as exercise)
Anyway, back to emoji.

We have gadgets, but we still need to chain them!
3. WRITING OUR EMOJI CHAIN
POKÉDEX OF RISC-V EMOJI GADGETS

- >4000 instructions available in total
- Logic and, add, sub, ...
- Conditional and unconditional branches (forward & backward)
- Many registers: a1357, s1359, t0246, gp, ra (but no sp 😞)
- Very few immediates
- Lots of floating-point instructions
POKÉDEX OF RISC-V EMOJI GADGETS

- >4000 instructions available in total
- Logic and, add, sub, ...
- Conditional and unconditional branches (forward & backward)
- ... only one ‘CSR’ instruction: `csrrs ra,0xbf8,gp` 😐
- Many registers: a1357, s1359, t0246, gp, ra (but no sp 😶)
- Very few immediates
- Lots of floating-point instructions

A tiny bit of everything... looks like a yard sale!
POKÉDEX OF RISC-V EMOJI GADGETS

- >4000 instructions available in total
- Logic and, add, sub, ...
- Conditional and unconditional branches (forward & backward)
- ... only one ‘CSR’ instruction: `csrrs ra,0xbf8,gp`

A tiny bit of everything... looks like a yard sale!

- Many registers: a1357, s1359, t0246, gp, ra (but no sp 😞)
- Very few immediates
- Lots of floating-point instructions

A tiny bit of everything... looks like a yard sale!
FROM BASIC BLOCKS TO ARBITRARY CODE EXECUTION

Stage 1

init

Unpacker $U$ (embedded $P_{enc}$)
FROM BASIC BLOCKS TO ARBITRARY CODE EXECUTION

Stage 1

init

Unpacker $U$ (embedded $P_{enc}$)

Payload (unpacked by $U$)
Stage 1

init

Unpacker $U$ (embedded $P_{enc}$)

Payload (unpacked by $U$)
FROM BASIC BLOCKS TO ARBITRARY CODE
EXECUTION

Stage 1

<table>
<thead>
<tr>
<th>init</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unpacker ( \mathcal{U} )</strong> (embedded ( \mathcal{P}_{\text{enc}} ))</td>
</tr>
</tbody>
</table>

Payload (unpacked by \( \mathcal{U} \))

Gadgets:

\[
\begin{align*}
  &a_1++ \\
  &*(\text{byte}*)a_3 = a_1 \\
  &a_3++
\end{align*}
\]
ARBITRARY PAYLOAD ENCODING

Initial Payload:

0x03
0x20
0x10
Initial Payload: 0x03 0x20 0x10

Encoded $P_{\text{enc}}$: 
*a1++; a1++; a1++;*(byte*)a3 = a1; a3++
a1++ … $\times (0x20 - 0x03) = 0x1D$ *(byte*)a3 = a1; a3++
a1++ … $\times (0x10 - 0x20) = 0xFO$ *(byte*)a3 = a1

Decoded Payload: 0x03 0x20 0x10
### ARBITRARY PAYLOAD ENCODING

<table>
<thead>
<tr>
<th>Initial Payload:</th>
<th>Encoded $P_{\text{enc}}$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x03</td>
<td>a1++; a1++; a1++;</td>
</tr>
<tr>
<td>0x20</td>
<td></td>
</tr>
<tr>
<td>0x10</td>
<td></td>
</tr>
<tr>
<td>Initial Payload:</td>
<td>Encoded $P_{enc}$:</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>0x03</td>
<td>$a1++; a1++; a1++;$</td>
</tr>
<tr>
<td></td>
<td>$<em>(\text{byte}</em>)a3 = a1; a3++$</td>
</tr>
<tr>
<td>0x20</td>
<td></td>
</tr>
<tr>
<td>0x10</td>
<td></td>
</tr>
</tbody>
</table>
Initial Payload:  \( \mathcal{P}_{\text{enc}}: \)

- **0x03**: 
  - \( a1++; \ a1++; \ a1++; \)
  - \( *(\text{byte}*)a3 = a1; \ a3++ \)

- **0x20**: 
  - \( a1++... \times (0x20 - 0x03) = 0x1D \)

- **0x10**
## Arbitrary Payload Encoding

<table>
<thead>
<tr>
<th>Initial Payload:</th>
<th>Encoded $\mathcal{P}_{enc}$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x03</td>
<td><code>a1++; a1++; a1++;</code></td>
</tr>
<tr>
<td></td>
<td><code>*(\text{byte*})a3 = a1; a3++</code></td>
</tr>
<tr>
<td>0x20</td>
<td><code>a1++... \times (0x20 - 0x03) = 0x1D</code></td>
</tr>
<tr>
<td></td>
<td><code>*(\text{byte*})a3 = a1; a3++</code></td>
</tr>
<tr>
<td>0x10</td>
<td></td>
</tr>
</tbody>
</table>

---

*emoji Shellcoding: 🚀, 🌟, and 🌟 @ DEFCON*
ARBITRARY PAYLOAD ENCODING

Initial Payload: 0x03 0x20 0x10

Encoded $P_{\text{enc}}$:

- $a1++; a1++; a1++$;
- $*(\text{byte}*)a3 = a1; a3++$
- $a1++ \ldots \times (0x20 - 0x03) = 0x1D$
- $*(\text{byte}*)a3 = a1; a3++$
- $a1++ \ldots \times (0x10 - 0x20) = 0xF0$
- $*(\text{byte}*)a3 = a1$

Decoded Payload: 0x03 0x20 0x10
**Initial Payload:**

0x03

0x20

0x10

**Encoded Payload:**

\[
P_{\text{enc}}: \quad a1++; \, a1++; \, a1++;
\]

\[
*(\text{byte}*)a3 = a1; \, a3++
\]

\[
a1++ \times (0x20 - 0x03) = 0x1D
\]

\[
*(\text{byte}*)a3 = a1; \, a3++
\]

\[
a1++ \times (0x10 - 0x20) = 0xF0
\]

\[
*(\text{byte}*)a3 = a1
\]

**Decoded Payload:**

0x03

0x20

0x10
<table>
<thead>
<tr>
<th>Initial Payload:</th>
<th>Encoded $\mathcal{P}_{\text{enc}}$:</th>
<th>Decoded Payload:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x03</td>
<td>$a_1++; a_1++; a_1++;$ \n$<em>(\text{byte}</em>)a_3 = a_1; a_3++$</td>
<td>0x03</td>
</tr>
<tr>
<td>0x20</td>
<td>$a_1++ \times (0x20 - 0x03) = 0x1D$ \n$<em>(\text{byte}</em>)a_3 = a_1; a_3++$</td>
<td></td>
</tr>
<tr>
<td>0x10</td>
<td>$a_1++ \times (0x10 - 0x20) = 0xF0$ \n$<em>(\text{byte}</em>)a_3 = a_1$</td>
<td></td>
</tr>
</tbody>
</table>
### ARBITRARY PAYLOAD ENCODING

<table>
<thead>
<tr>
<th>Initial Payload:</th>
<th>Encoded $P_{\text{enc}}$:</th>
<th>Decoded Payload:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x03</td>
<td>a1++; a1++; a1++;</td>
<td>0x03</td>
</tr>
<tr>
<td></td>
<td><em>(byte</em>)a3 = a1; a3++</td>
<td></td>
</tr>
<tr>
<td>0x20</td>
<td>a1++; ... × (0x20 − 0x03) = 0x1D</td>
<td>0x20</td>
</tr>
<tr>
<td></td>
<td><em>(byte</em>)a3 = a1; a3++</td>
<td></td>
</tr>
<tr>
<td>0x10</td>
<td>a1++; ... × (0x10 − 0x20) = 0xFF</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>(byte</em>)a3 = a1</td>
<td></td>
</tr>
</tbody>
</table>
## Arbitrary Payload Encoding

<table>
<thead>
<tr>
<th>Initial Payload</th>
<th>Encoded $P_{\text{enc}}$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x03</td>
<td>$a_1++; a_1++; a_1++;</td>
</tr>
<tr>
<td></td>
<td>$(\text{byte}*)a_3 = a_1; a_3++$</td>
</tr>
<tr>
<td>0x20</td>
<td>$a_1++; \ldots \times (0x20 - 0x03) = 0x1D$</td>
</tr>
<tr>
<td></td>
<td>$(\text{byte}*)a_3 = a_1; a_3++$</td>
</tr>
<tr>
<td>0x10</td>
<td>$a_1++; \ldots \times (0x10 - 0x20) = 0xF0$</td>
</tr>
<tr>
<td></td>
<td>$(\text{byte}*)a_3 = a_1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decoded Payload</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x03</td>
</tr>
<tr>
<td>0x20</td>
</tr>
<tr>
<td>0x10</td>
</tr>
</tbody>
</table>
## ARBITRARY PAYLOAD ENCODING

<table>
<thead>
<tr>
<th>Initial Payload:</th>
<th>Encoded $P_{\text{enc}}$:</th>
<th>Decoded Payload:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x03</td>
<td><code>a1++; a1++; a1++;</code></td>
<td>0x03</td>
</tr>
<tr>
<td></td>
<td><code>*(byte*)a3 = a1; a3++</code></td>
<td></td>
</tr>
<tr>
<td>0x20</td>
<td><code>a1++ ... \times (0x20 - 0x03) = 0x1D</code></td>
<td>0x20</td>
</tr>
<tr>
<td></td>
<td><code>*(byte*)a3 = a1; a3++</code></td>
<td></td>
</tr>
<tr>
<td>0x10</td>
<td><code>a1++ ... \times (0x10 - 0x20) = 0xF0</code></td>
<td>0x10</td>
</tr>
<tr>
<td></td>
<td><code>*(byte*)a3 = a1</code></td>
<td></td>
</tr>
<tr>
<td>e2 99</td>
<td>8c c2</td>
<td>a9 ef</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>0x0</td>
<td>0x2</td>
<td>0x4</td>
</tr>
</tbody>
</table>

GDB OVER BEAMER TIME!
GDB OVER BEAMER TIME!

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e2</td>
<td>99</td>
<td>8c</td>
<td>c2</td>
<td>a9</td>
<td>ef</td>
<td>b8</td>
</tr>
<tr>
<td>0x0</td>
<td>0x2</td>
<td>0x4</td>
<td>0x6</td>
<td>0x8</td>
<td>...</td>
<td>0x62</td>
</tr>
</tbody>
</table>

```
add s3, s3, s8
sw a1, 0(a3)
bnez a5, +0x5a
...
add  s3, s3, s8
bnez a3, +0x2a
add  a3, a3, t2
```

```plaintext
*(byte*)a3 = a1
```

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>0000</td>
<td>00AB</td>
</tr>
<tr>
<td>a3</td>
<td>0000</td>
<td>8000</td>
</tr>
<tr>
<td>t2</td>
<td>0000</td>
<td>0001</td>
</tr>
<tr>
<td>a5</td>
<td>B100</td>
<td>D5AC</td>
</tr>
<tr>
<td>s3</td>
<td>530F</td>
<td>25F8</td>
</tr>
<tr>
<td>s8</td>
<td>03C4</td>
<td>9ECC</td>
</tr>
</tbody>
</table>

```
... 00 ...
```

```
...
```

```
... 0x8000 0x8001
```
```
add s3, s3, s8  
sw a1, 0(a3)  
bnez a5, +0x5a
... 
add s3, s3, s8  
bnez a3, +0x2  
add a3, a3, t2

*(byte*)a3 = a1
```

### GDB over Beamer Time!

#### Assembly Code:

```
add    s3, s3, s8  \{ *(byte*)a3 = a1 \}
sw     a1, 0(a3)  \{ *(byte*)a3 = a1 \}
```

```
bnez   a5, +0x5a
...
add    s3, s3, s8  \{ a3++ \}
bnez   a3, +0x2
add    a3, a3, t2
```

### Memory Addresses:

<table>
<thead>
<tr>
<th>Address</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0</td>
<td>0x0000 00AB</td>
</tr>
<tr>
<td>0x2</td>
<td>0x0000 8000</td>
</tr>
<tr>
<td>0x4</td>
<td>0x0000 0001</td>
</tr>
<tr>
<td>0x6</td>
<td>B100 D5AC</td>
</tr>
<tr>
<td>0x68</td>
<td>56D3 C4C4</td>
</tr>
<tr>
<td>0x6A</td>
<td>03C4 9ECC</td>
</tr>
<tr>
<td>0x62</td>
<td>A1</td>
</tr>
<tr>
<td>0x64</td>
<td>AB</td>
</tr>
<tr>
<td>0x66</td>
<td>0x8000</td>
</tr>
<tr>
<td>0x8001</td>
<td></td>
</tr>
</tbody>
</table>
### GDB OVER BEAMER TIME!

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a2</strong></td>
<td><strong>©</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>e2</strong></td>
<td><strong>9d</strong></td>
<td><strong>8c</strong></td>
<td><strong>e2</strong></td>
<td><strong>99</strong></td>
</tr>
<tr>
<td><strong>8c</strong></td>
<td><strong>c2</strong></td>
<td><strong>a9</strong></td>
<td><strong>ef</strong></td>
<td><strong>b8</strong></td>
<td><strong>8f</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x00</td>
<td>0x02</td>
<td>0x04</td>
<td>0x06</td>
<td>0x08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0x62</td>
<td>0x64</td>
</tr>
</tbody>
</table>

```
add s3, s3, s8
sw a1, 0(a3)
bnez a5, +0x5a

... add s3, s3, s8
bnez a3, +0x2
add a3, a3, t2
```

```
*(byte*)a3 = a1
```

```
a1 0000 00AB
a3 0000 8000
t2 0000 0001
a5 B100 D5AC
s3 56D3 C4C4
s8 03C4 9ECC
```

```
... AB ...
```

```
... 0x8000 0x8001
```
```plaintext
add s3, s3, s8
sw a1, 0(a3)  \( \text{*(byte*)a3 = a1} \)
bnez a5, +0x5a

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{e2} & \text{99} & \text{8c} & \text{c2} & \text{a9} & \text{ef} \\
\hline
0x0 & 0x2 & 0x4 & 0x6 & 0x8 & \cdots \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|}
\hline
\text{e2} & \text{9d} & \text{8c} & \text{e2} \\
\hline
0x62 & 0x64 & 0x66 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{add} & \text{s3}, \text{s3}, \text{s8} & \text{a3}++ \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{a1} & 0000 & 00AB \\
\hline
\text{a3} & 0000 & 8000 \\
\hline
\text{t2} & 0000 & 0001 \\
\hline
\text{a5} & \text{B100 D5AC} & \text{5A98 6390} \\
\hline
\text{s3} & \text{03C4 9ECC} & \text{0000 0000} \\
\hline
\text{s8} & \text{0x8000} & \text{0x8001} \\
\hline
\end{array}
\]
```
**GDB OVER BEAMER TIME!**

```
<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e2 99</td>
<td>8c c2</td>
<td>a9 ef</td>
<td>b8 8f</td>
<td>...</td>
<td>e2 9d</td>
<td>8c e2</td>
<td>99 8b</td>
</tr>
<tr>
<td>0x0</td>
<td>0x2</td>
<td>0x4</td>
<td>0x6</td>
<td>0x8</td>
<td>...</td>
<td>0x62</td>
<td>0x64</td>
</tr>
</tbody>
</table>

### Code Snippet

```
```add s3, s3, s8
sw a1, 0(a3)
bnez a5, +0x5a
...```

```
```add s3, s3, s8
bnez a3, +0x2
add a3, a3, t2
```

### Instructions

- `add s3, s3, s8` *(byte*)
- `sw a1, 0(a3)`
- `bnez a5, +0x5a`
- `...`
- `add s3, s3, s8`
- `bnez a3, +0x2`
- `add a3, a3, t2`

### Memory

```
<table>
<thead>
<tr>
<th>Address</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0000</td>
<td>00AB</td>
</tr>
<tr>
<td>0x8000</td>
<td>0x8001</td>
</tr>
</tbody>
</table>
```

### Registers

- **a1**: 0000 00AB
- **a3**: 0000 8000
- **t2**: 0000 0001
- **a5**: B100 D5AC
- **s3**: 5A98 6390
- **s8**: 03C4 9ECC

### Abbreviations

- **AB**: 0x8000 0x8001
GDB OVER BEAMER TIME!

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e2  99</td>
<td>8c c2</td>
<td>a9 ef</td>
<td>b8 8f</td>
<td>...</td>
<td>e2 9d</td>
<td>8c e2</td>
</tr>
<tr>
<td>0x0</td>
<td>0x2</td>
<td>0x4</td>
<td>0x6</td>
<td>0x8</td>
<td>...</td>
<td>0x62</td>
</tr>
</tbody>
</table>

```
add    s3,s3,s8  
sw     a1,0(a3)  
bnez   a5,+0x5a  
...    
add    s3,s3,s8  
bnez   a3,+0x2   
add    a3,a3,t2  
```

```
*(byte*)a3 = a1
```

```
a1 | 0000 00AB
a3 | 0000 8001
a5 | B100 D5AC
s3 | 5A98 6390
s8 | 03C4 9ECC
```

```
... | AB | ...
... | 0x8000 | 0x8001
```
```
GDB OVER BEAMER TIME!

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e2</td>
<td>99</td>
<td>8c</td>
<td>c2</td>
<td>a9</td>
<td>ef</td>
</tr>
<tr>
<td>b8</td>
<td>8f</td>
<td>...</td>
<td></td>
<td>e2</td>
<td>9d</td>
</tr>
<tr>
<td>0x0</td>
<td>0x2</td>
<td>0x4</td>
<td>0x6</td>
<td>0x8</td>
<td>...</td>
</tr>
</tbody>
</table>

```

```
add s3, s3, s8
sw a1, 0(a3)  *(byte*)a3 = a1
bnez a5, +0x5a
...
add s3, s3, s8
bnez a3, +0x2  a3++
add a3, a3, t2
```

```
| a1  | 0000 00AB |
| a3  | 0000 8001 |
| t2  | 0000 0001 |
| a5  | B100 D5AC |
| s3  | 5A98 6390 |
| s8  | 03C4 9ECC |

```

```
<table>
<thead>
<tr>
<th></th>
<th>AB</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>0x8000</td>
<td>0x8001</td>
</tr>
</tbody>
</table>
```

emoji Shellcoding: 🔧, 💡, and 🎆 @ DEF CON
DEMO ON RV64GC SIFIVE UNLEASHED BOARD
FILLING THE VOIDS

Just solve a variant of subset sum (RTFM Wikipedia).

- Trivial if we use only 3 and 4-byte emojis...
- ... harder if we want to minimize number of emojis

⇒ dynamic programming (takes 4'13" to implement)

Fantastic, we get polymorphism for free!
Just solve a variant of subset sum (RTFM Wikipedia).
Just solve a variant of subset sum (RTFM Wikipedia).

- Trivial if we use only 3 and 4-byte emojis...
FILLING THE VOIDS

- Just solve a variant of subset sum (RTFM Wikipedia).
  - Trivial if we use only 3 and 4-byte emojis ...
  - ... harder if we want to minimize number of emojis
    ⇒ dynamic programming (takes 4’13” to implement)
Just solve a variant of subset sum (RTFM Wikipedia).

- Trivial if we use only 3 and 4-byte emojis ...
- ... harder if we want to minimize number of emojis
  ⇒ dynamic programming (takes 4’13” to implement)

Fantastic, we get polymorphism for free!
### Just solve a variant of subset sum (RTFM Wikipedia).
- Trivial if we use only 3 and 4-byte emojis ...
- ... harder if we want to minimize number of emojis
  ⇒ dynamic programming (takes 4’13” to implement)

### Fantastic, we get polymorphism for free!
### FILLING THE VOIDS

<table>
<thead>
<tr>
<th>emoji</th>
<th>0x0</th>
<th>0x2</th>
<th>0x4</th>
<th>0x6</th>
<th>0x8</th>
<th>...</th>
<th>0x62</th>
<th>0x64</th>
<th>0x66</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑</td>
<td>e2 99</td>
<td>8c c2</td>
<td>a9 ef</td>
<td>b8 8f</td>
<td>...</td>
<td>e2 9d</td>
<td>8c e2</td>
<td>99 8b</td>
<td></td>
</tr>
</tbody>
</table>

- Just solve a variant of subset sum (RTFM Wikipedia).
  - Trivial if we use only 3 and 4-byte emojis ...
  - ... harder if we want to minimize number of emojis
    ⇒ dynamic programming (takes 4’13” to implement)

- Fantastic, we get polymorphism for free!
Just solve a variant of subset sum (RTFM Wikipedia).

- Trivial if we use only 3 and 4-byte emojis ...
- ... harder if we want to minimize number of emojis
  \[ \Rightarrow \text{dynamic programming (takes 4'13" to implement)} \]

Fantastic, we get polymorphism for free!
Can we do our nop-sled with the 📦 emoji?
Can we do our nop-sled with the 🎤 emoji?
TRIVIA

Can we do our nop-sled with the 🚴 emoji?

✅ ©
Can we do our nop-sled with the 🧵 emoji?
Can we do our nop-sled with the 🛑 emoji?

This is executable code!
DEMO ON RV32 ESPRESSIF ESP32-C3 BOARD
666. CONCLUSION
CONCLUSION
ACKNOWLEDGMENTS

The following were harmed during the making of this project:

- LuaLatex (segfault)
- TexStudio (memory leaks, crashes, ...)
- Acrobat Reader (major glitches)
- Various PDF readers (glitches)
- VLC (glitches on demo.mkv)
- Firefox (broke PDF.js persistently)
- xfce4-terminal (utter slowness)
- Noto Emoji (render issues)
- CUPS (crash)
- gcc (segfault)
- Windows (BSOD when compiling slides)

Emoji support is hard.
We went all the way to design a new code-reuse technique just to hack you with unsolicited “I ❤️ U” texts.

Code and documentation on:

https://github.com/RischardV/emoji-shellcoding

Short link: bit.do/dc30

Your friendly neighbourhood hackers

hadrien.barral@ens.fr   georges-axel.jaloyan@ens.fr