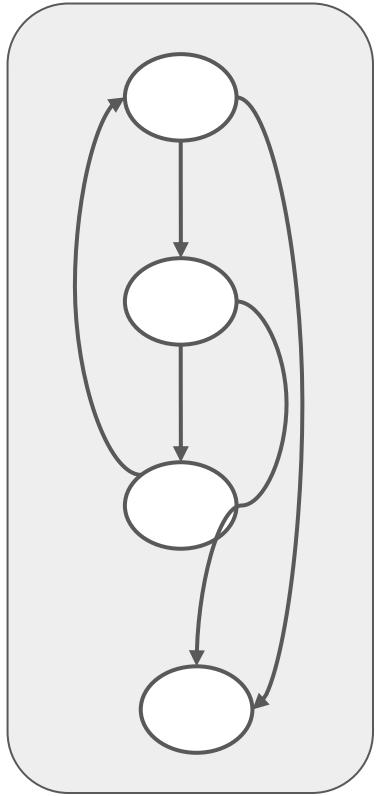
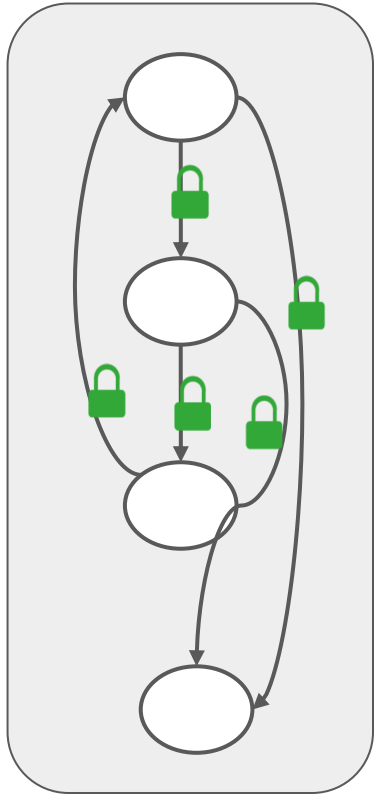
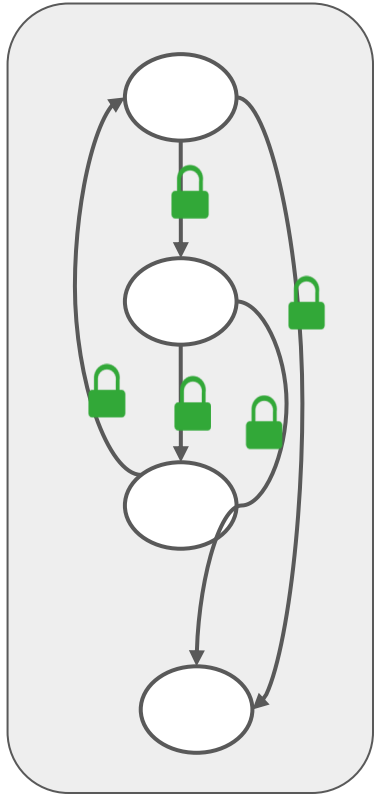


Practical Data-Only Attack Generation

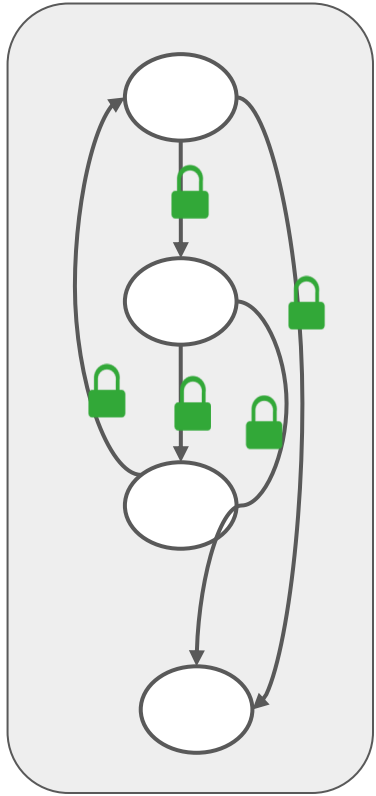
Brian Johannesmeyer, Asia Slowinska,
Herbert Bos, Cristiano Giuffrida

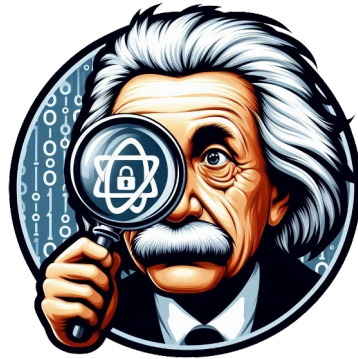
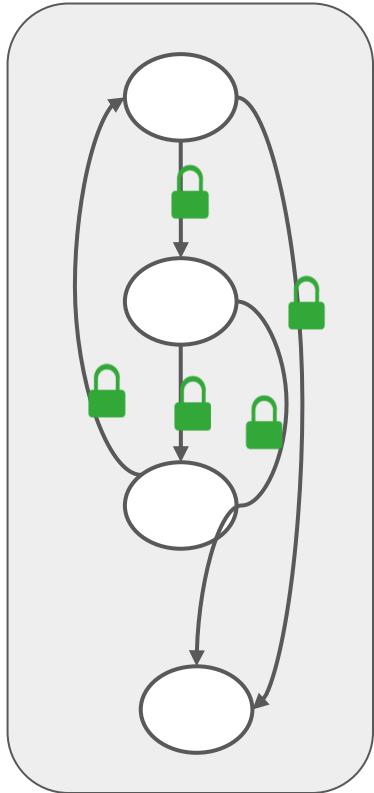


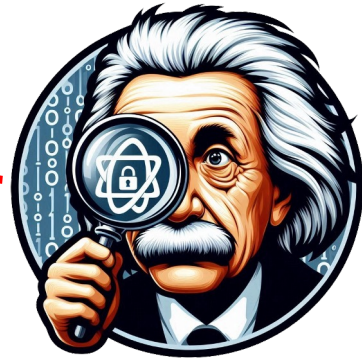
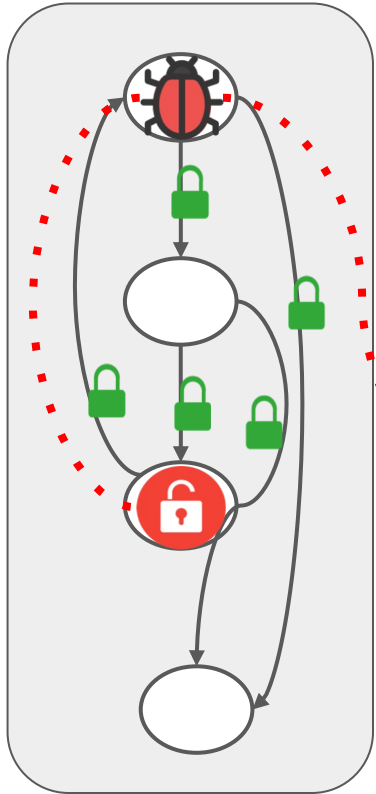


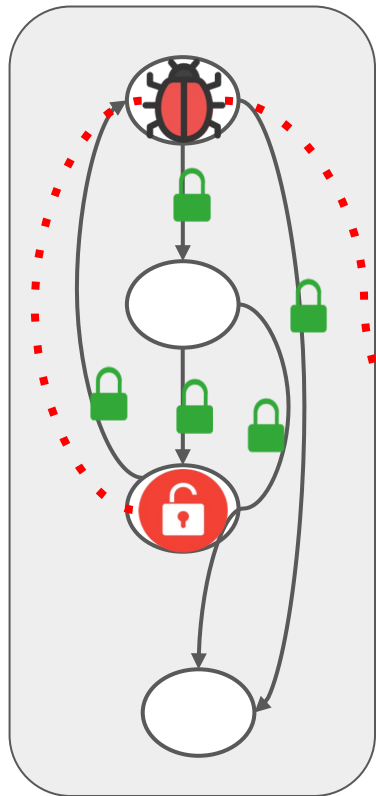


Data-only attacks!

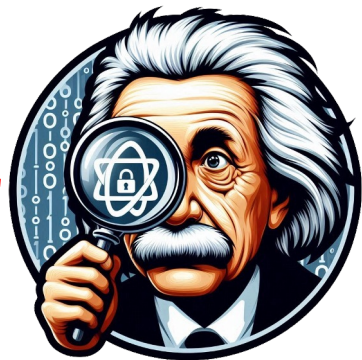




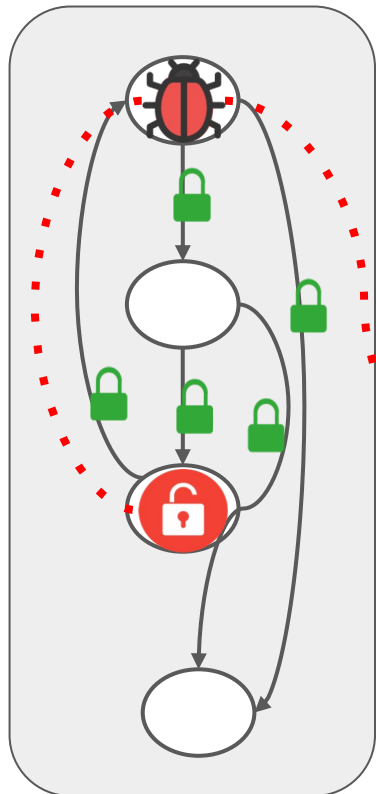




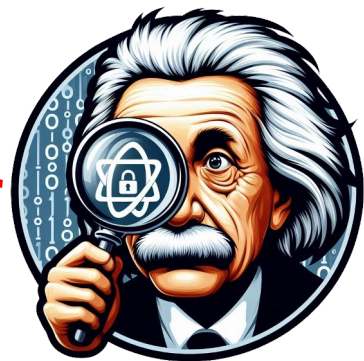
Data-only attacks!



*Builds **hundreds of exploits** against popular web servers*



Data-only attacks!

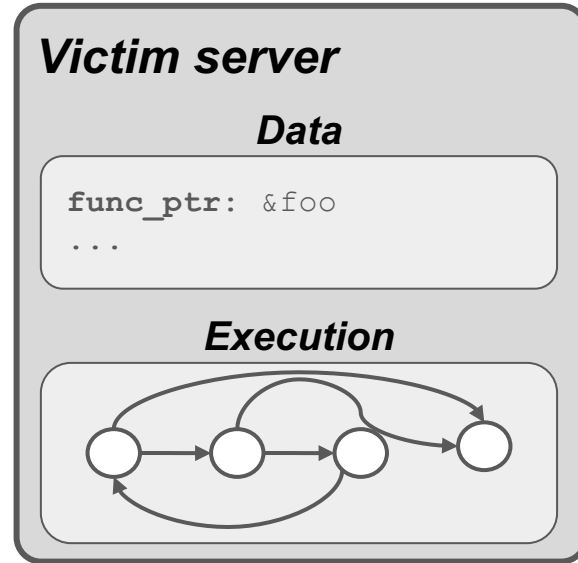


*Builds **hundreds of exploits** against popular web servers*

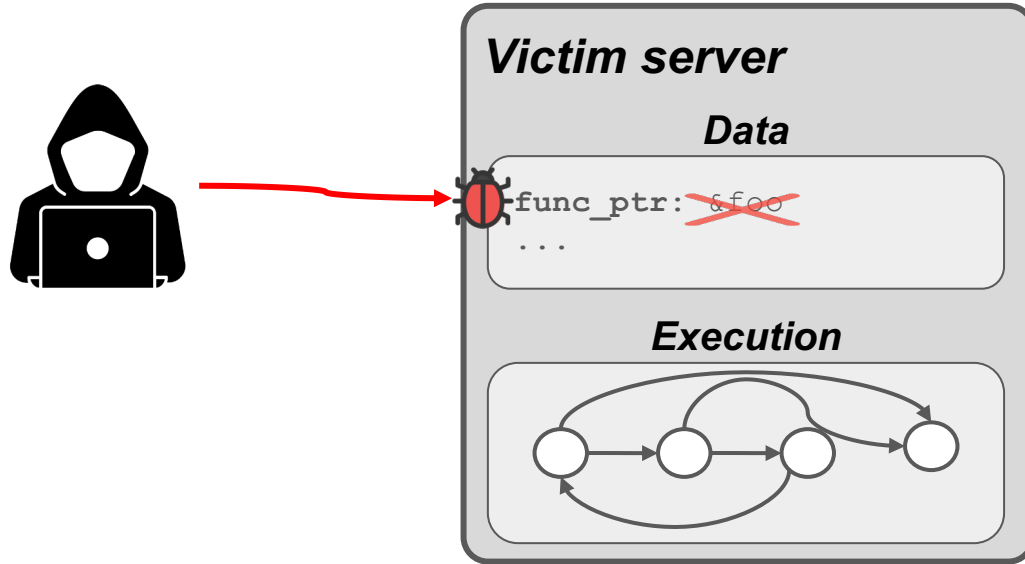


*Calls upon researchers and vendors to **rethink mitigation strategies***

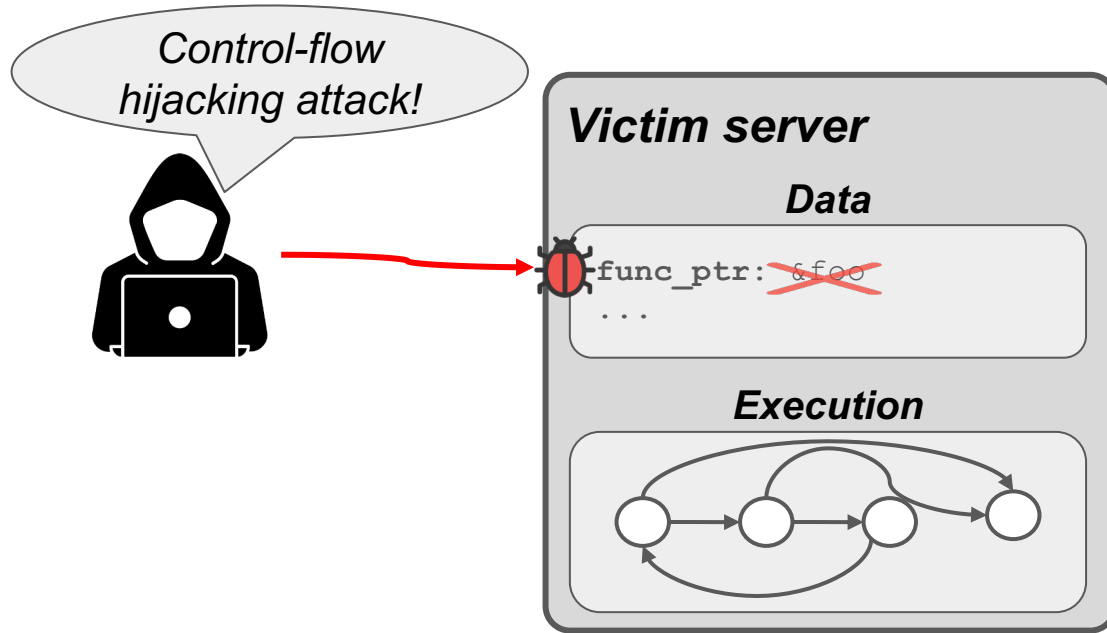
What is a data-only attack?



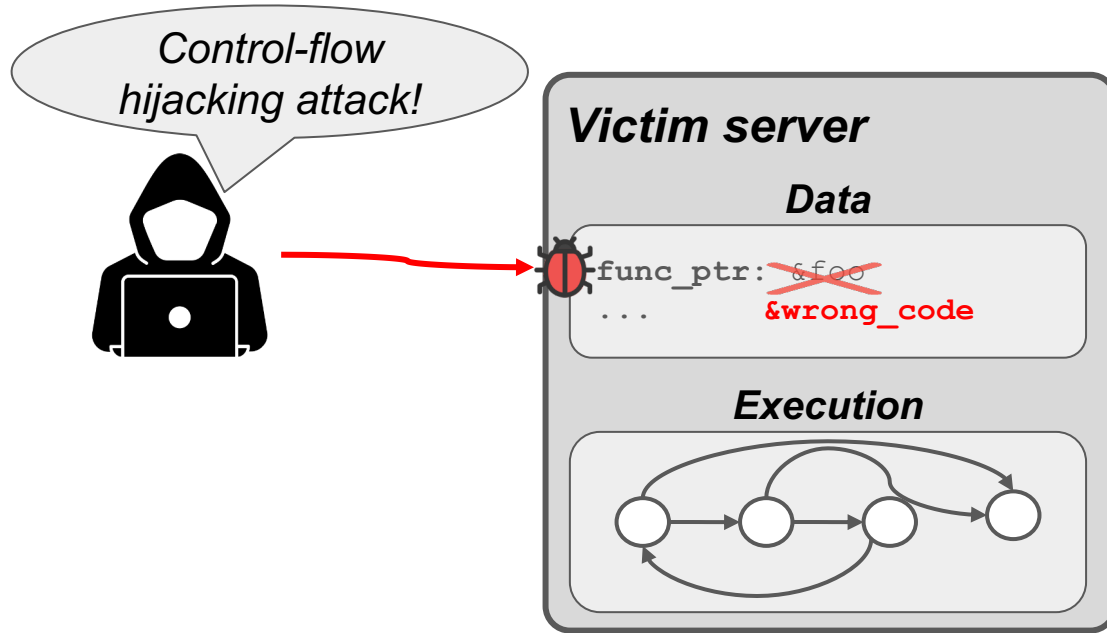
What is a data-only attack?



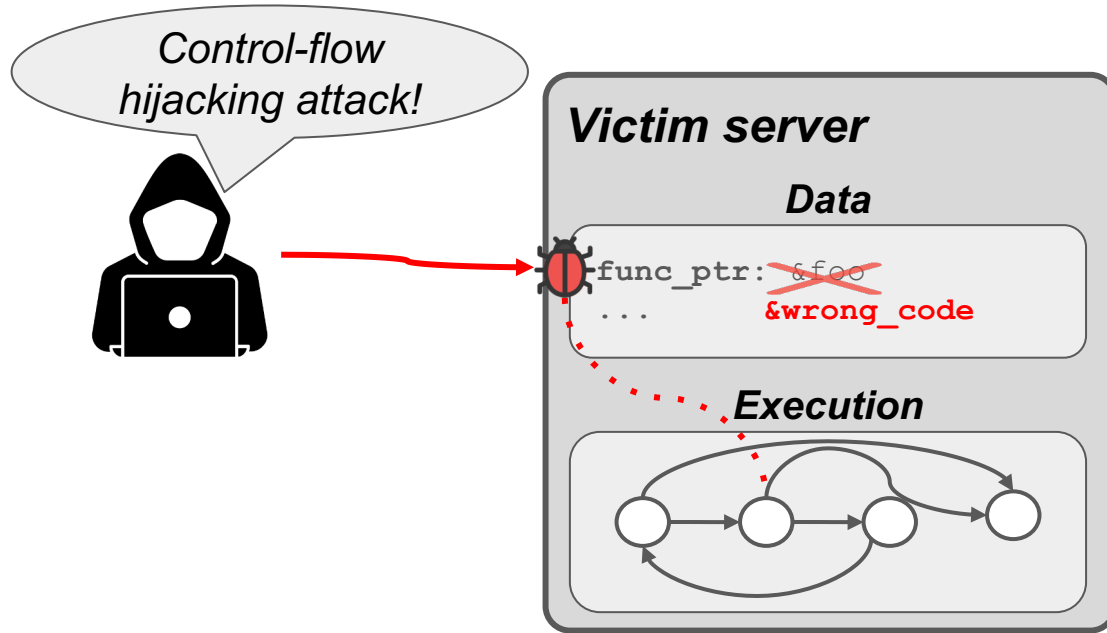
What is a data-only attack?



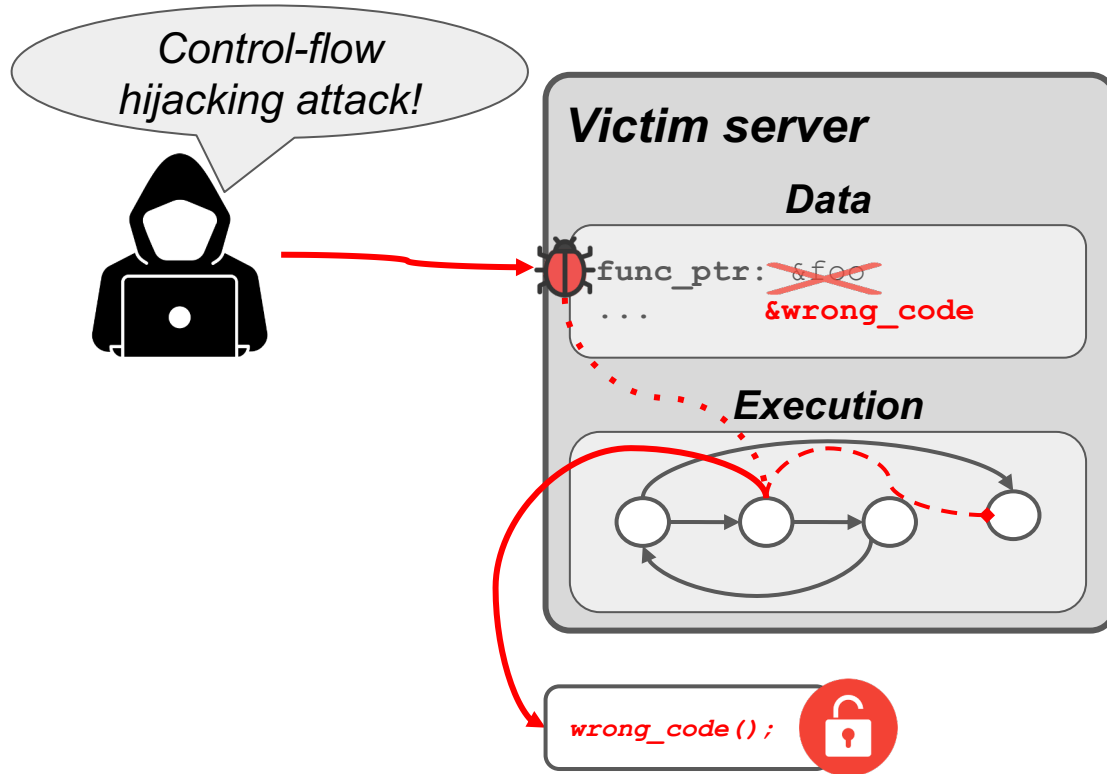
What is a data-only attack?



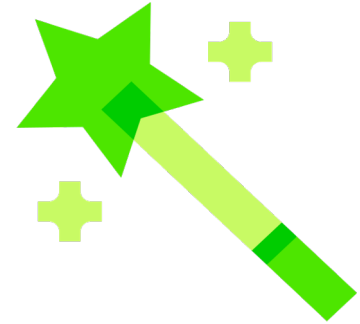
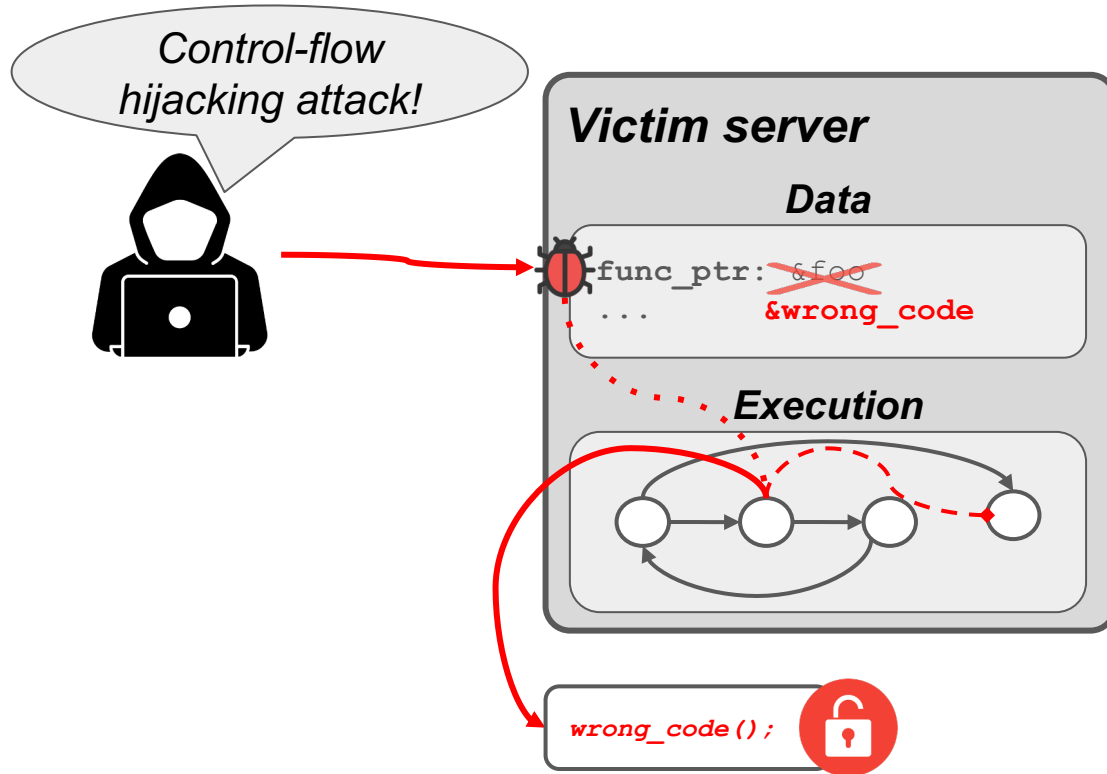
What is a data-only attack?



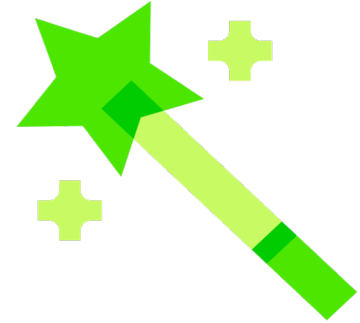
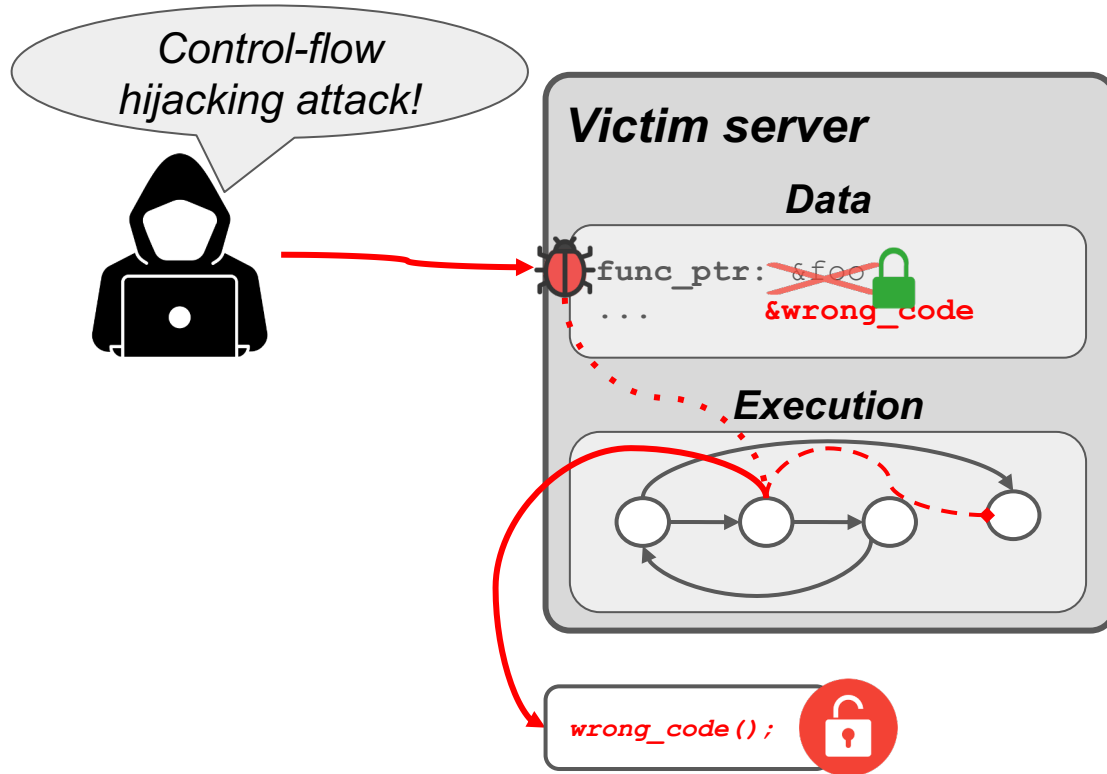
What is a data-only attack?



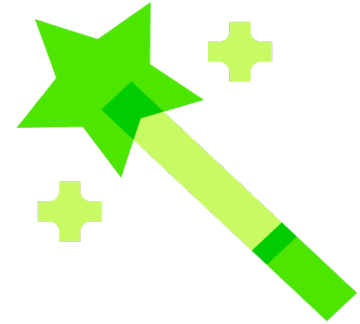
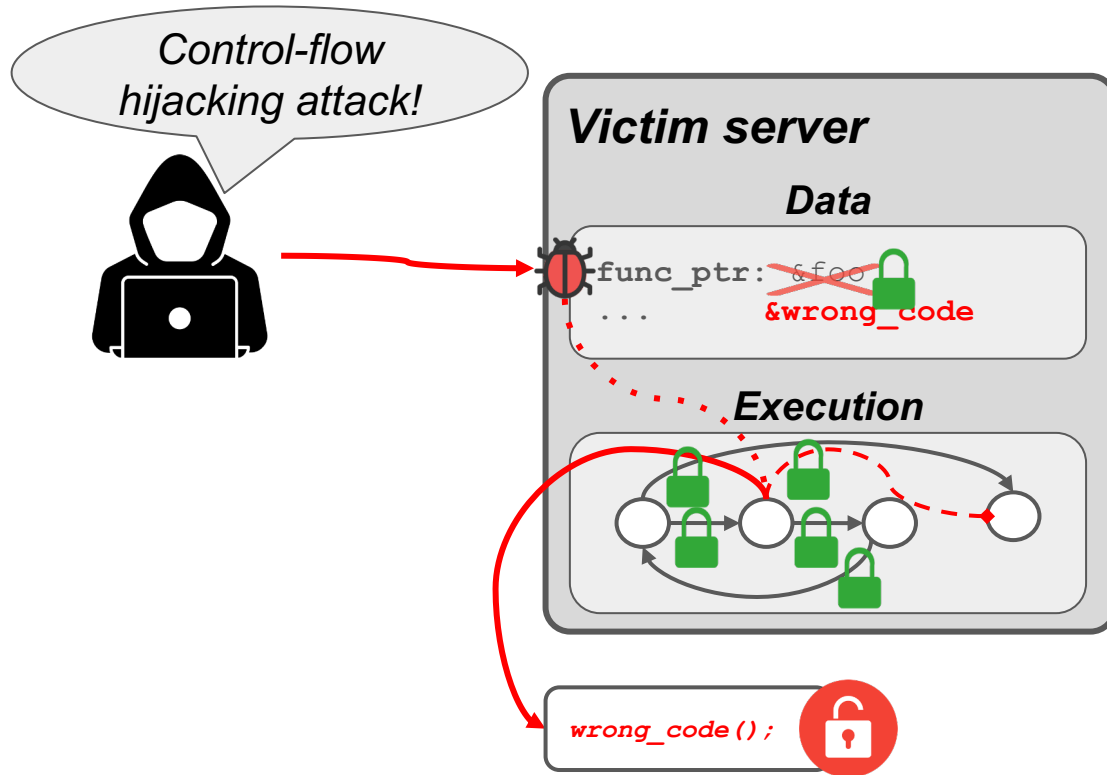
What is a data-only attack?



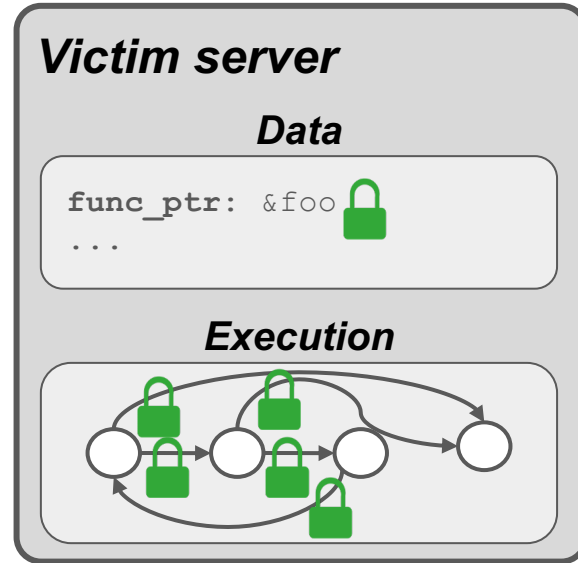
What is a data-only attack?



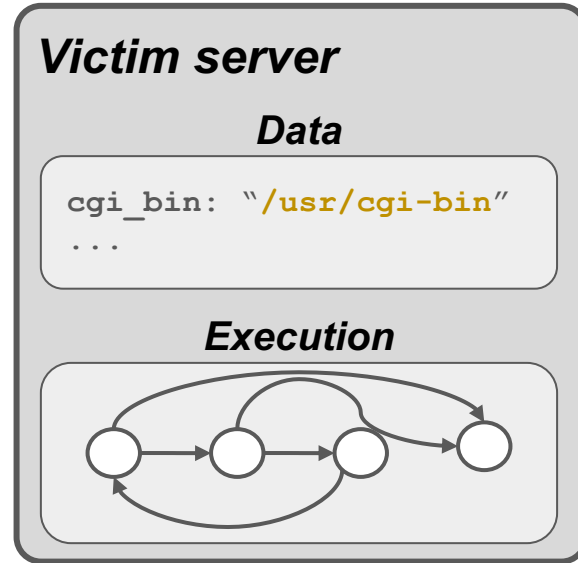
What is a data-only attack?



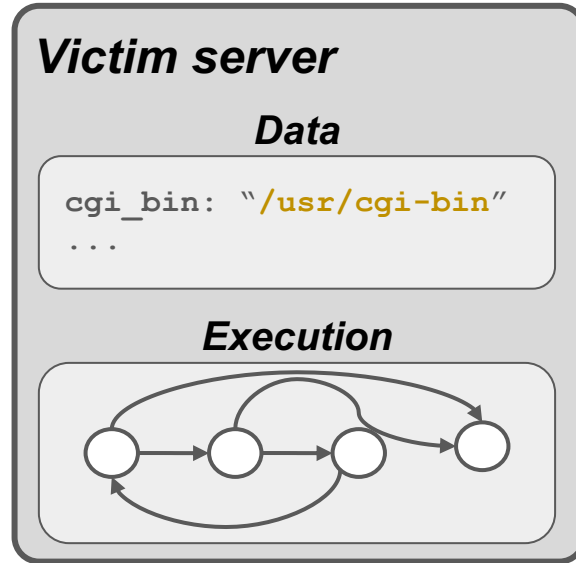
What is a data-only attack?



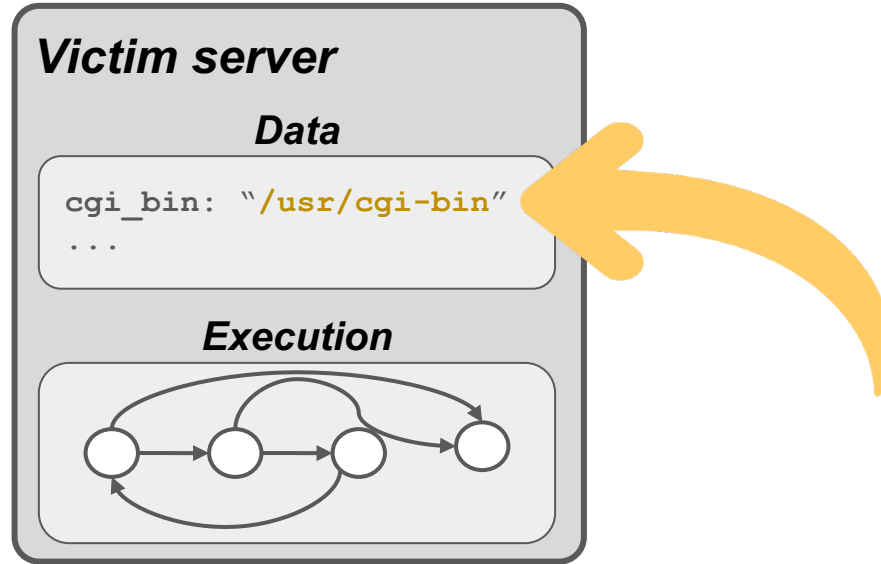
What is a data-only attack?



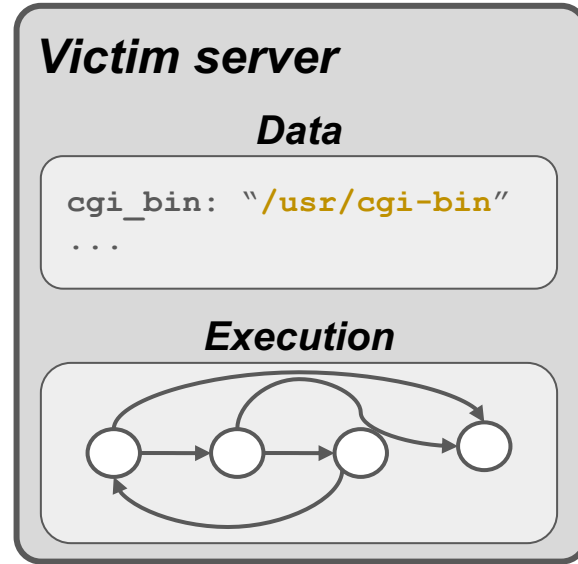
What is a data-only attack?



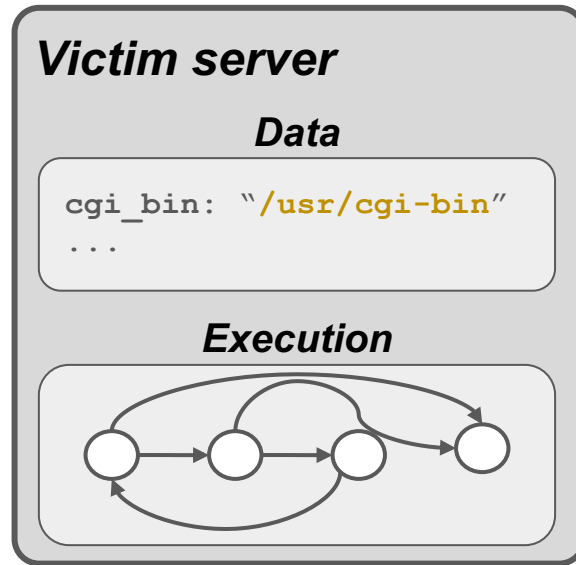
What is a data-only attack?



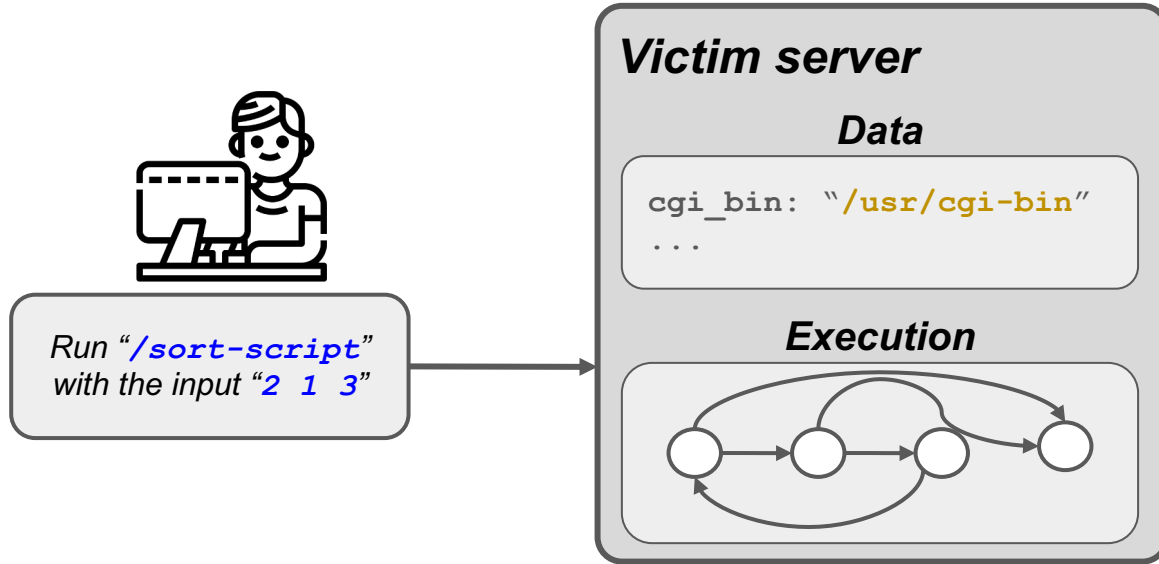
What is a data-only attack?



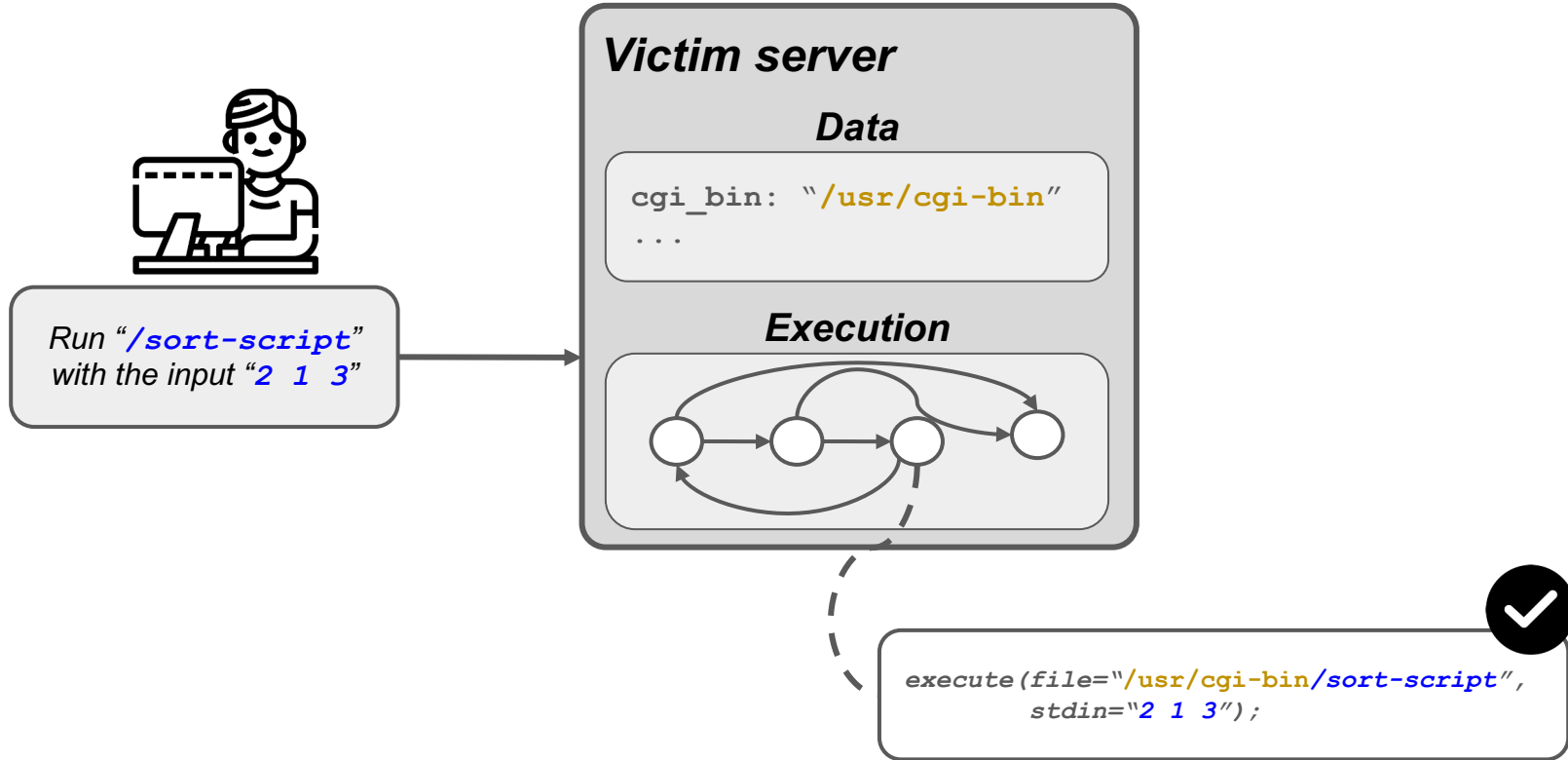
What is a data-only attack?



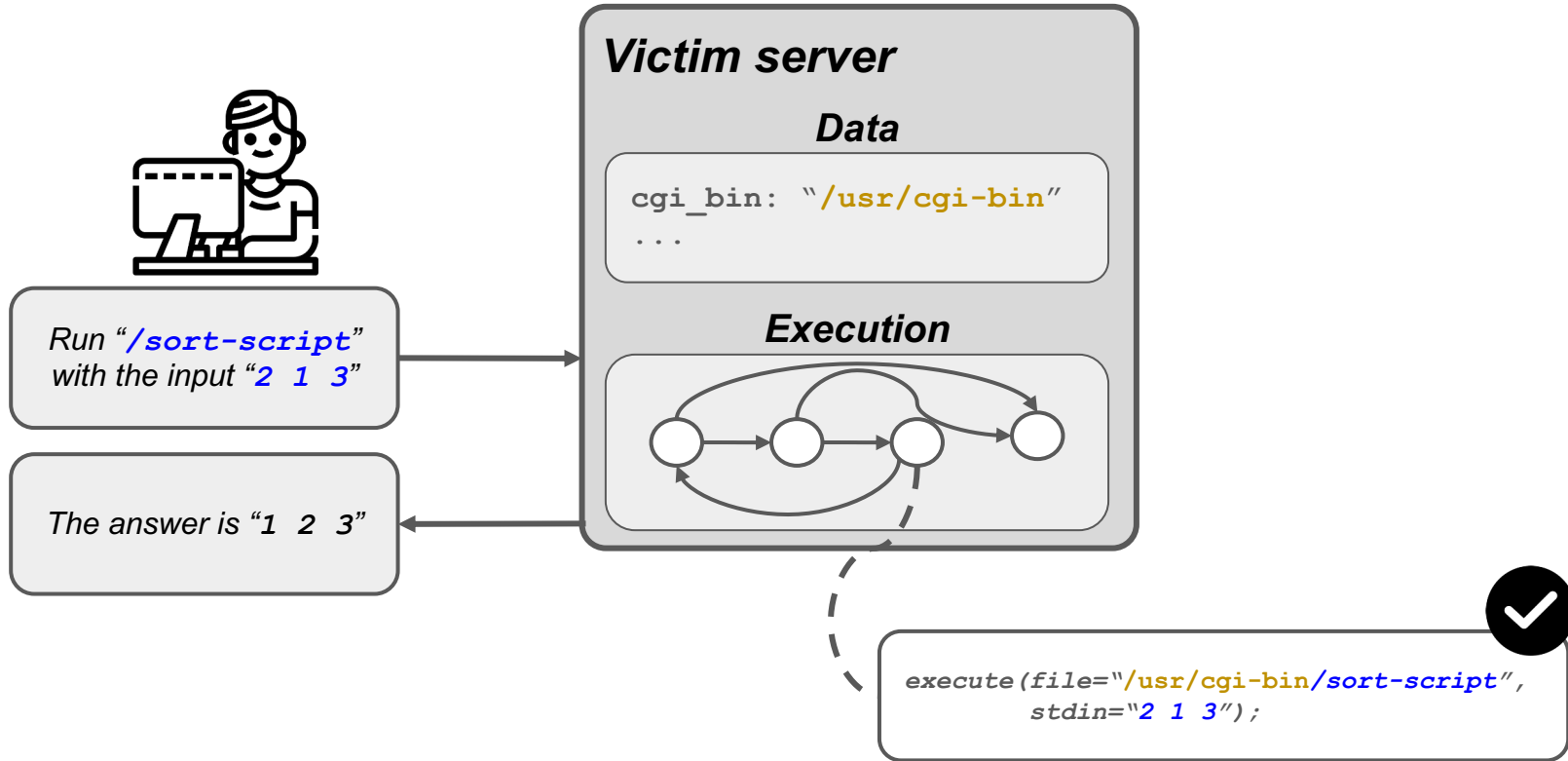
What is a data-only attack?



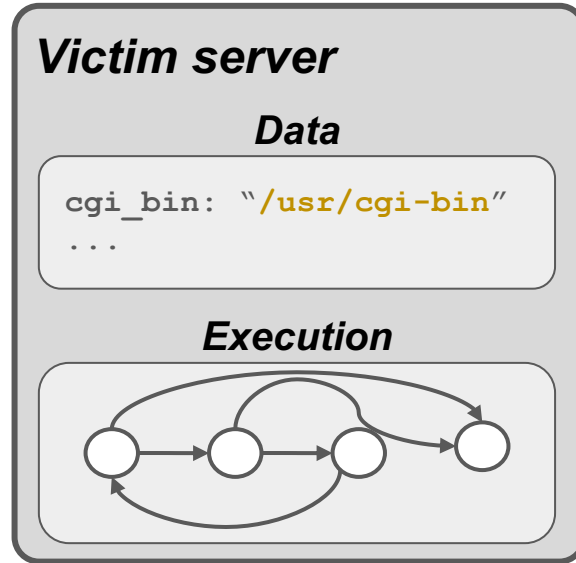
What is a data-only attack?



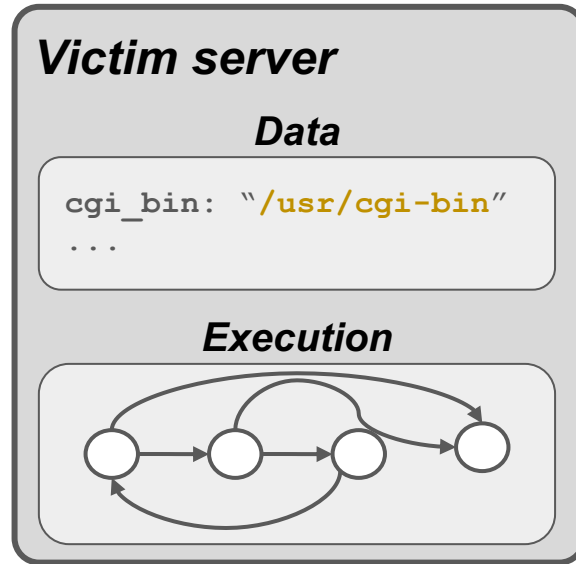
What is a data-only attack?



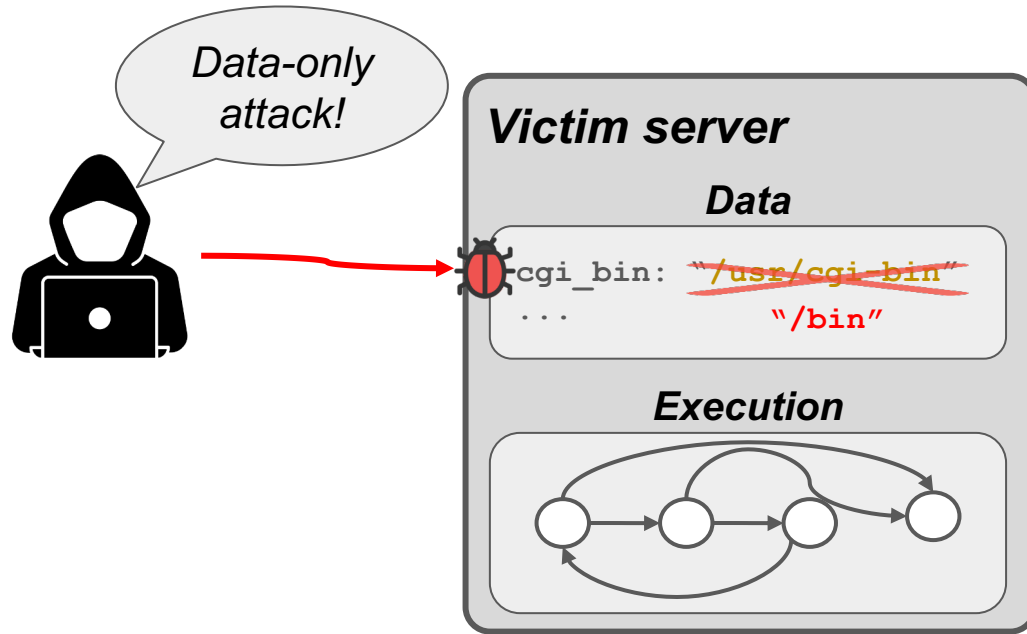
What is a data-only attack?



What is a data-only attack?



What is a data-only attack?

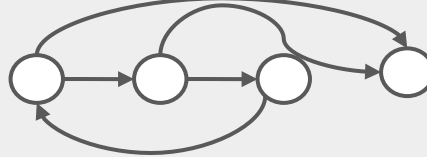


Victim server

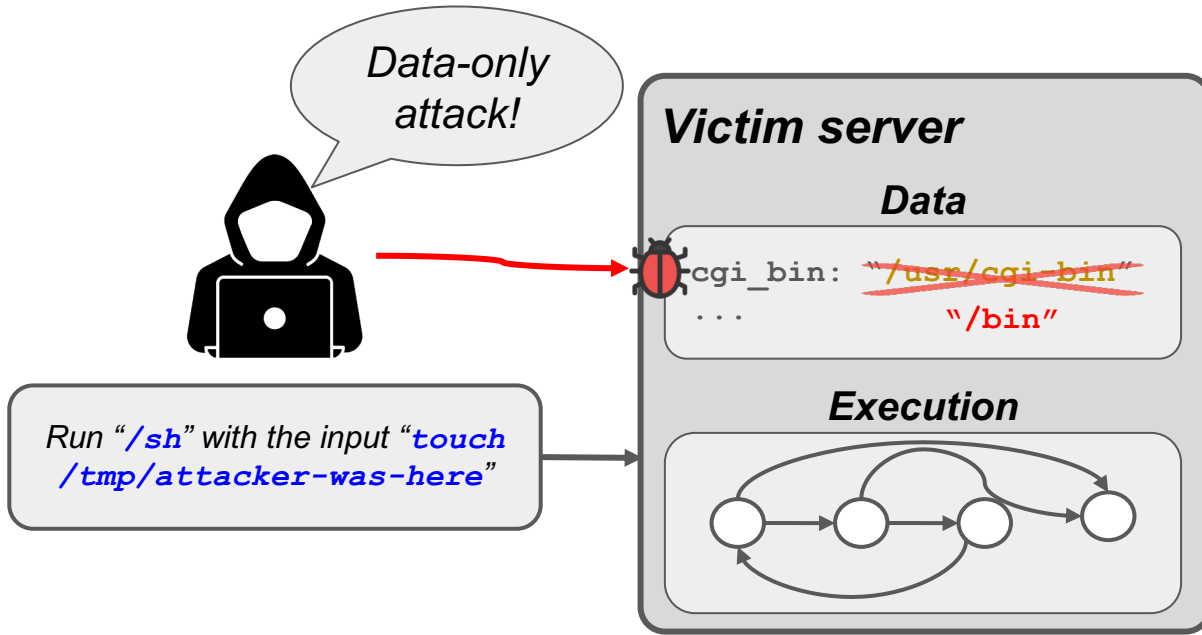
Data

```
cgi_bin: "/usr/cgi_bin"  
... "/bin"
```

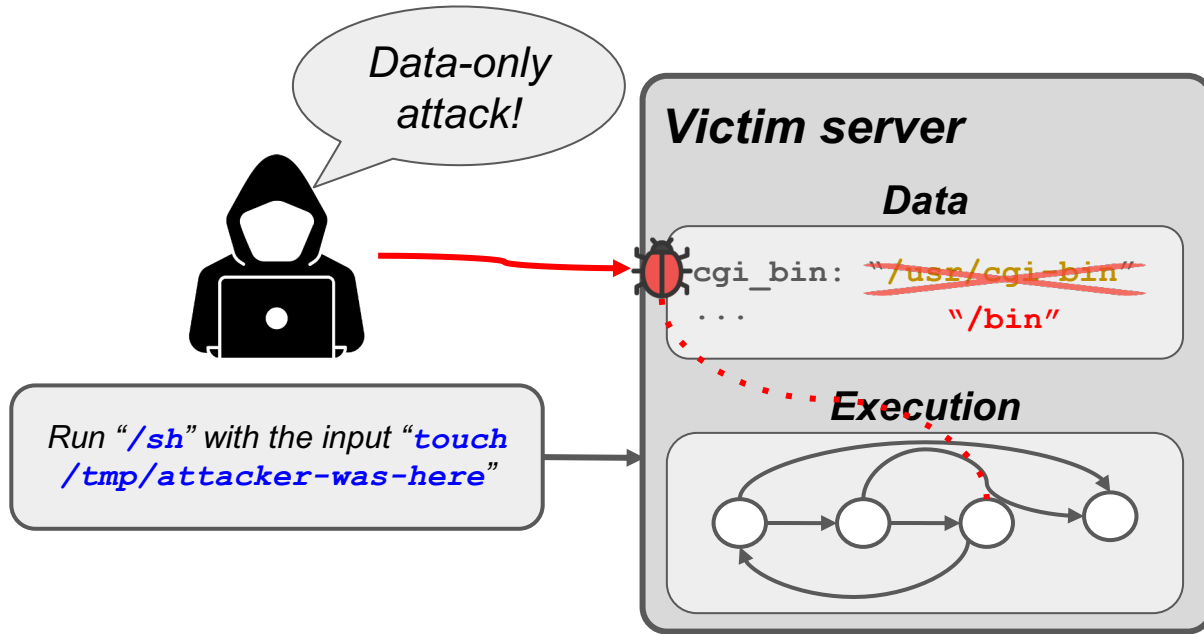
Execution



What is a data-only attack?



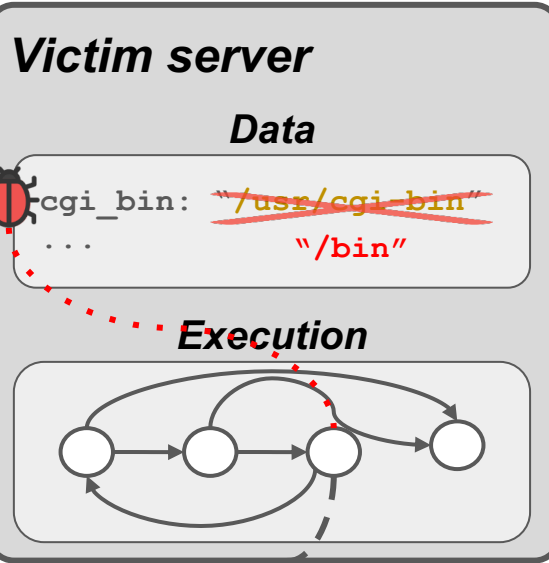
What is a data-only attack?



What is a data-only attack?



Data-only
attack!



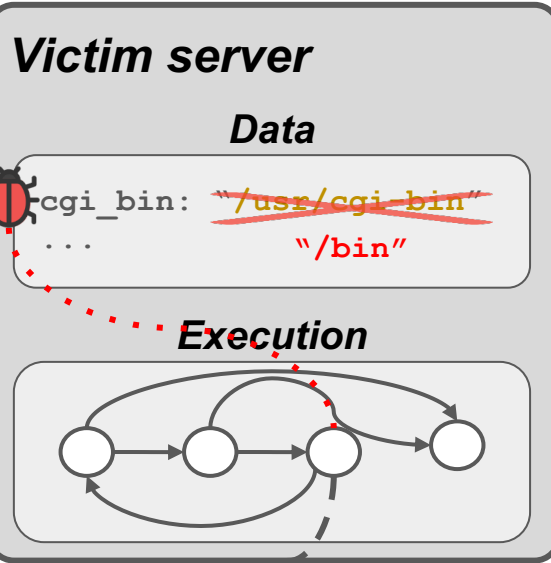
Run `"/sh"` with the input `touch /tmp/attacker-was-here`

```
execute(file="/bin/sh",  
        stdin="touch /tmp/attacker-was-here");
```

What is a data-only attack?



Data-only
attack!



Run `"/sh"` with the input `touch /tmp/attacker-was-here`

```
execute(file="/bin/sh",  
        stdin="touch /tmp/attacker-was-here");
```



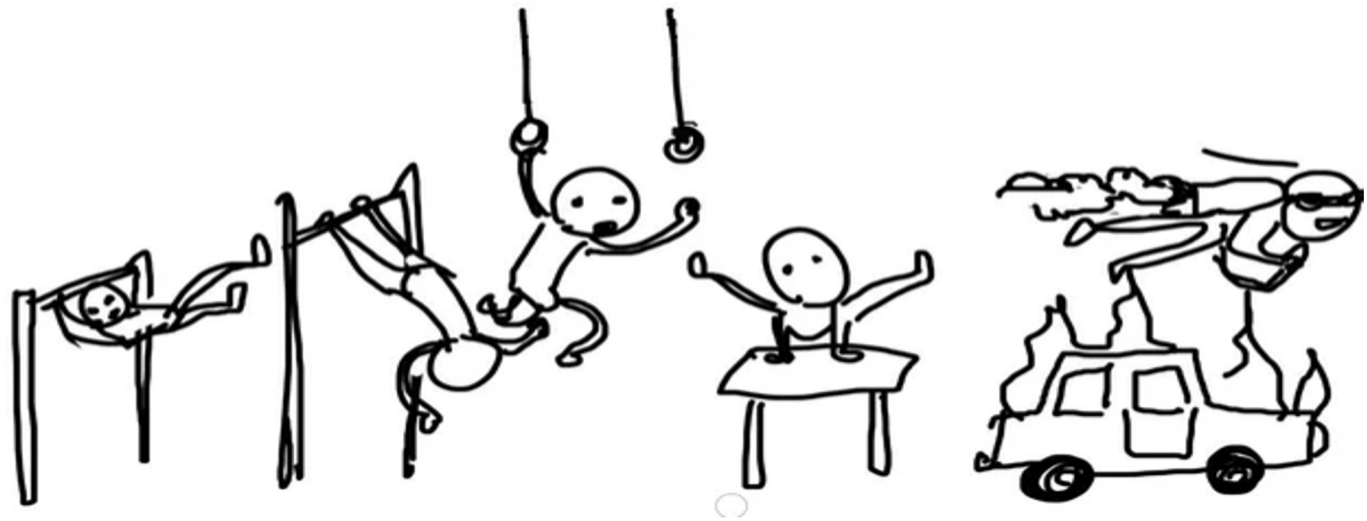
The attack does not corrupt
the *victim's control flow*.

The attack does not corrupt
the *victim's control flow*.

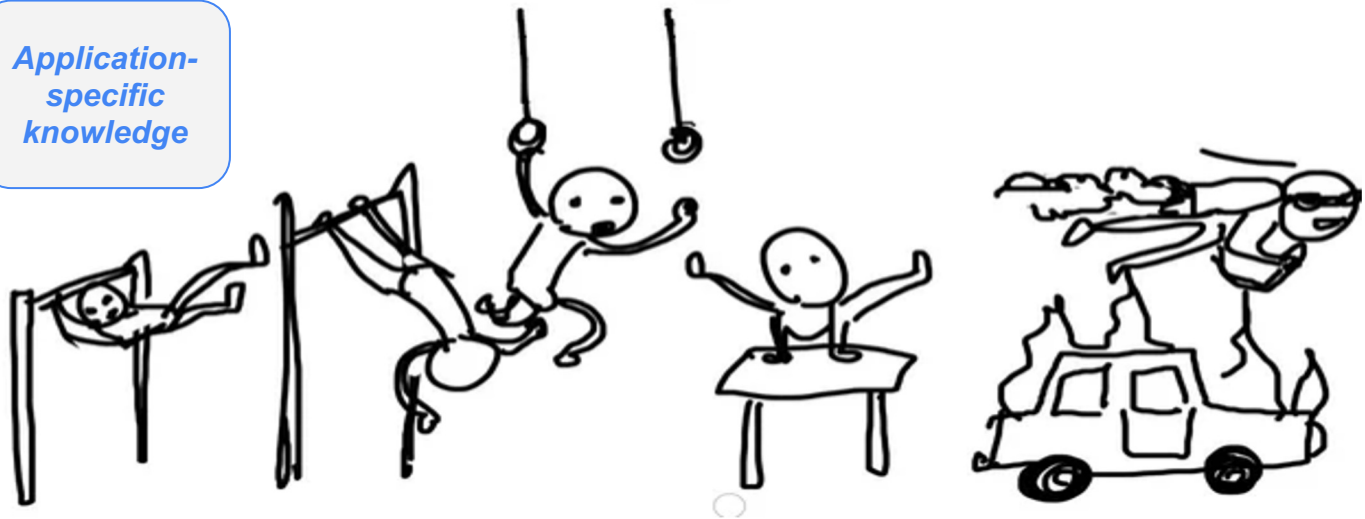
The attack does not corrupt
the *victim's control flow*.

It only corrupts ***its function arguments***.

Why data-only
attacks are
**considered
difficult**

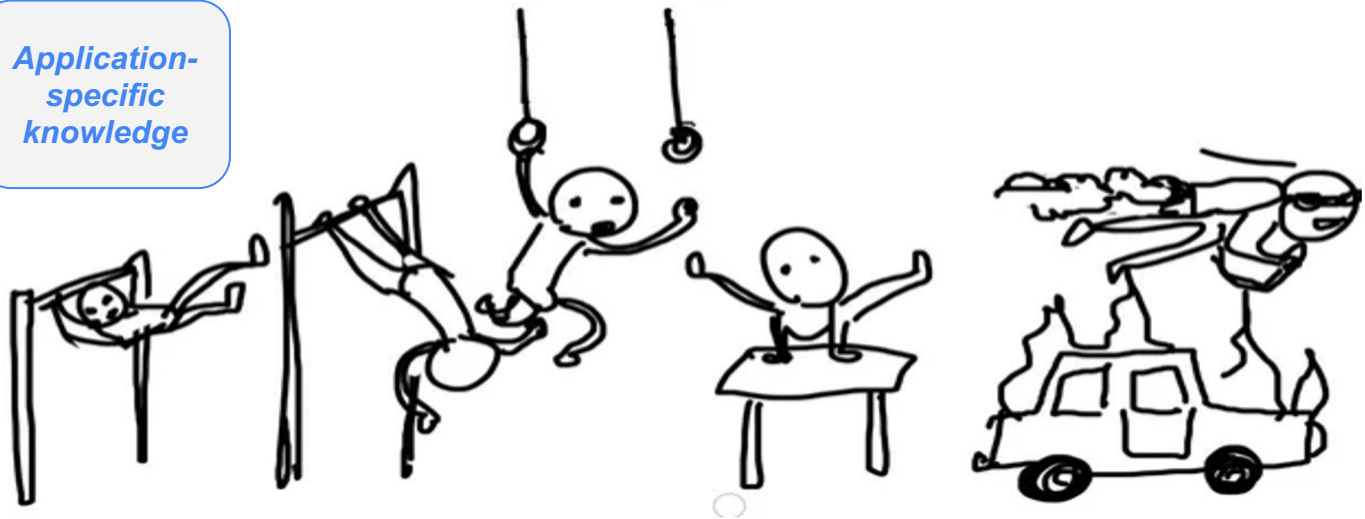


Application-specific knowledge



Why data-only attacks are **considered difficult**

Application-specific knowledge

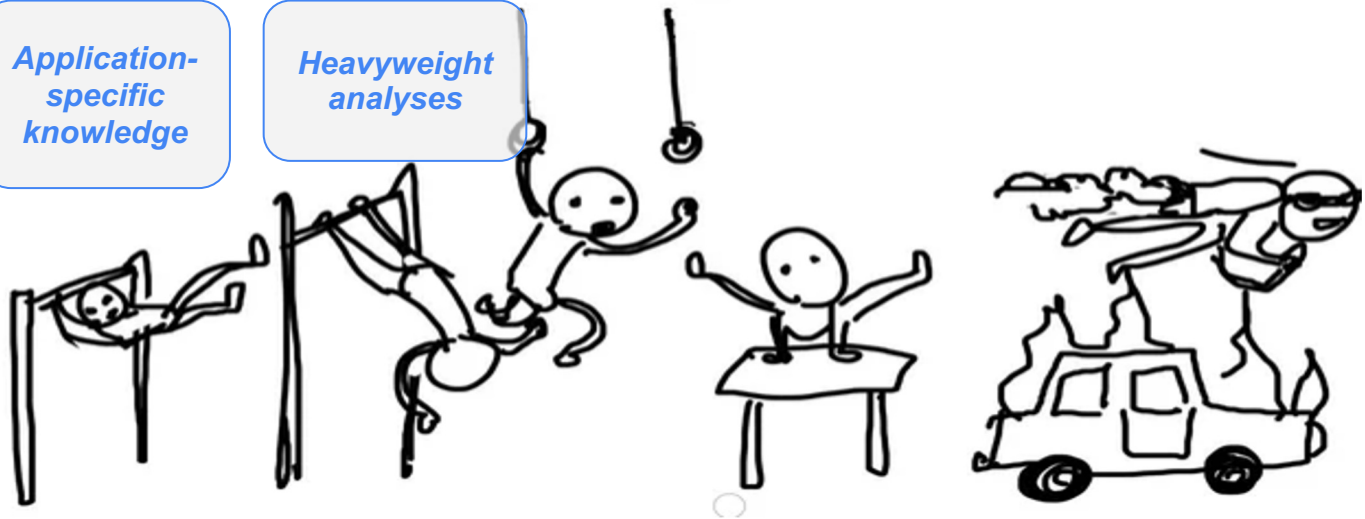


Why data-only attacks are **considered difficult**

Why data-only attacks are considered difficult

Application-specific knowledge

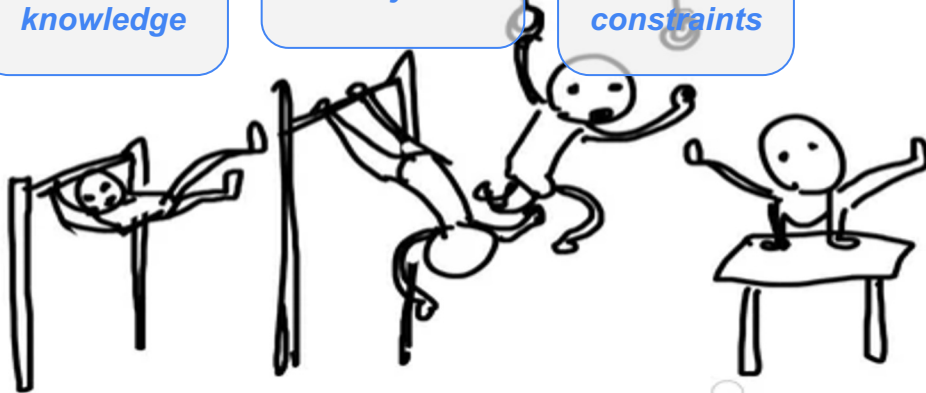
Heavyweight analyses



Application-specific knowledge

Heavyweight analyses

Complex dataflow constraints



Why data-only attacks are considered difficult

```
return!0<t?te3?crypt(-79,-13,a+crypt(-87,1-,crypt(-86,0,a1)+a):  
1,t<_2crypt(t+1,_,a):3,crypt(-94,-27+t,a)&&t==2?_<13?  
crypt(2,_,1,"%s %d %d\n"):9:16:t<0?te<72?crypt(,t,  
"@n'+,#'/*{}/w/#c&n/r,+}/k{*+/,w{&+/,#l+/,n(n+/,+##n+/,#  
;#q#n+/,+k#;+/,/r : 'd#*3,}{w+k w'Kc'+)e#;dq#l \\  
q#'+d'K#!/+k#;q#'}eKk{n}'/#;#q#n'}{#}w')}{n}'/+#n';d}nw' i;#\  
{n}'/n{#': r{#}'r nc{n}'/#{L,+ 'Kw' {r' iK;[{n}]'/w#q#n'wk mw' \  
i#wk{Kk{n}'/w{%'l#w#& i; :{n}'/*{q#'}ld;r'}{n}w'b/'&de}'c \  
;:{n}'-({}nw')/+,)##*}#nc,'#mw}'/+'d'+e)+;#rdq#w: nr' / ')}{r'l#}'n' '# \  
'+'+##(!'!"):t<-50?_==a?putchar(31[a]):crypt(-65,_,a+1):crypt((#a==')'+t,_,a+1)  
:0<t?crypt(2,2,"%s"):#a==/'/'||crypt(0,crypt(-61,*,a  
"!ek;dc i@bK'(c)-(w)*#n+3#l,{}:}nwLoca-0m .vbvks,fxntdCeqhiry",a+1);
```

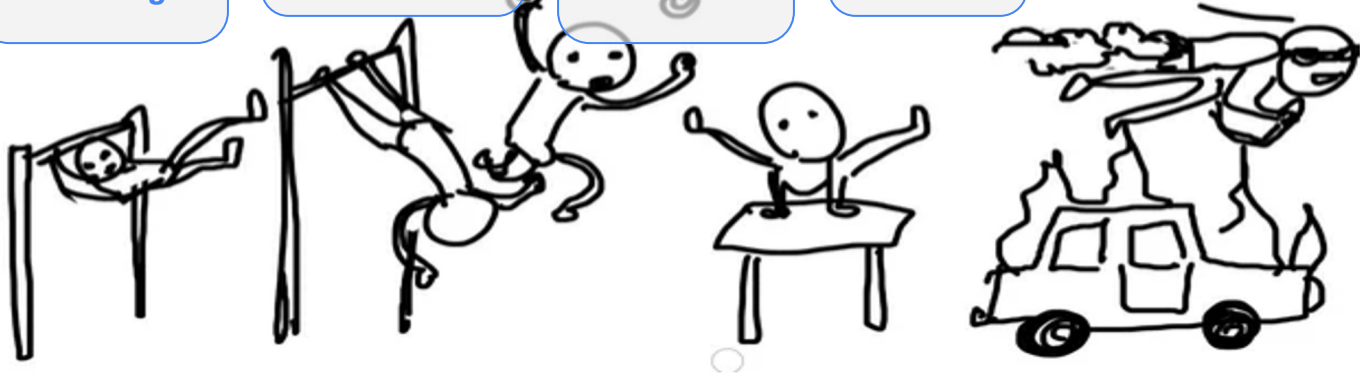


Application-specific knowledge

Heavyweight analyses

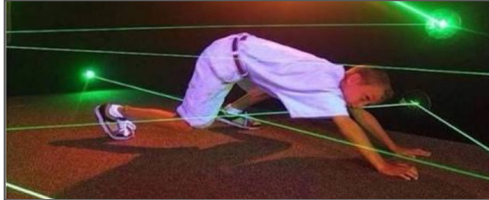
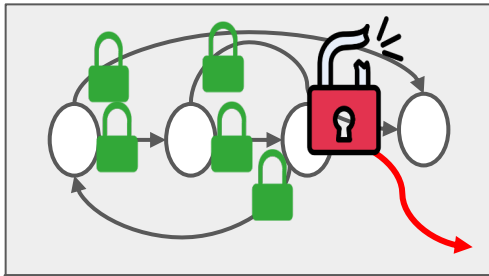
Complex dataflow constraints

Bypass defenses

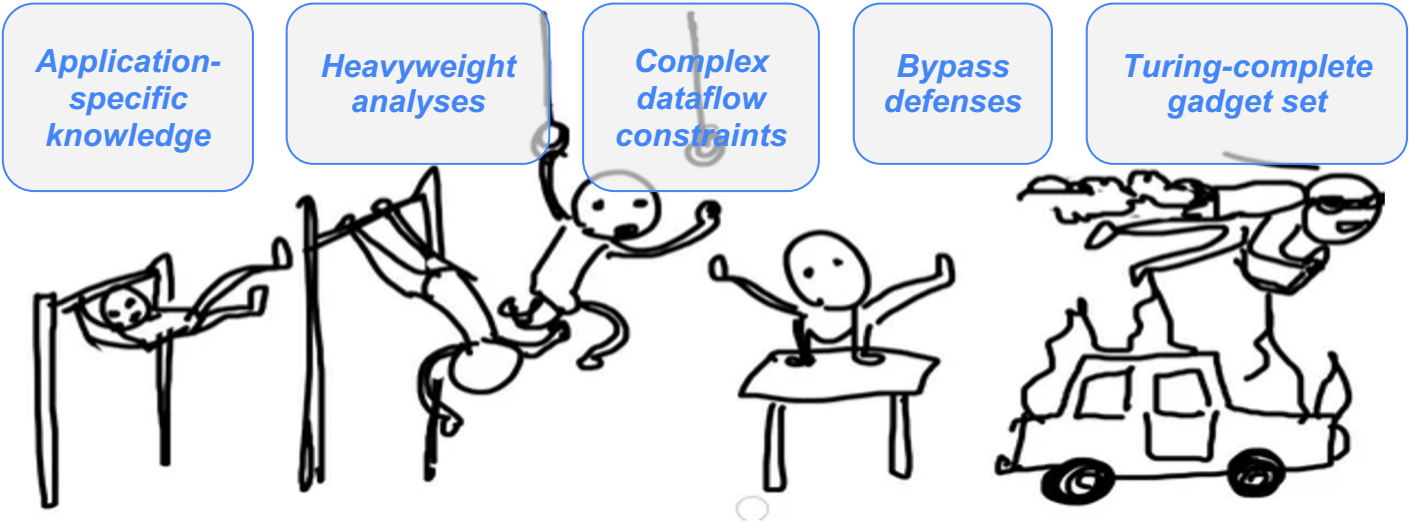


Why data-only attacks are considered difficult

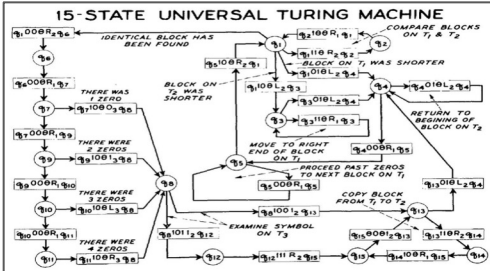
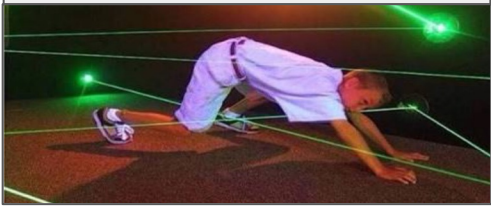
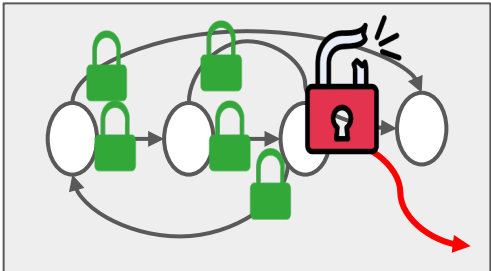
```
return!0<t?<t<3?crypt(-79,-13,a+crypt(-87,1-_,crypt(-86,0,a1)+a)):
1,t<_2crypt(t+1,_,a):3,crypt(-94,-27+t,a)&&t==22_<13?
crypt(2,_,1,"%s %d\n"):9:16:t<0?<-72?crypt(,t,
"@n'+,#'/*}{w/w#cdnr/+,}{r/*de}+,{/***/,w{**/,w#q#n+,#l+,/n(n+,/+n+,#\
#q#n+/,+k#;+/,/'r : 'd#*3,}{w+K w'K'+)e#;dq#'\ \
q#'+d'K#!/+k#;q#'}eK{#}w'r}eK{[n]'/#;#q#n'}{[n]}'/+#n';d}rw' i;#\
}[n]!}/n{#': r{#}'r nc{[n]}'#l,+K' {rw' iK;{[n]}'/w#q#n'wk mw' \
i#wk{K{[n]}'/w{#}'l#w#* i; :{n]}'/*{q#'}d;r'}{n(wb!/*de}'c \
;;{nL'-(}{rw}'+,)#'*)#nc,'#mw}'/+kd'+e+);#rdq#w! nr'/' ')}{r!#}'n' '# \
}'*+##(!'!'):t<-50?_==a?putchar(31[a]):crypt(-65,_,a+1):crypt({*a==}'/)+t,_,a+1)
:0<t?crypt(2,2,"%s"):+a==/'/'||crypt(0,crypt(-61,*a,
"!ek;dc i@bK'(q)-(w)*n+r3#l,{}:\mwLoca-0:m .vbvks,fxntdCeqhiry"),a+1);
```



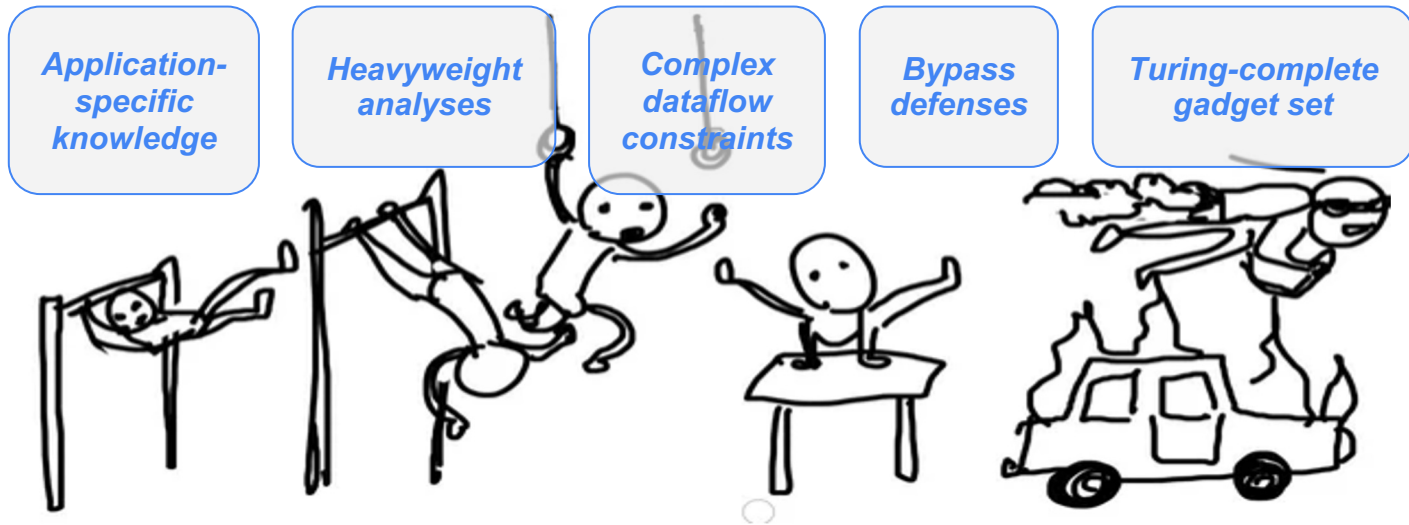
Why data-only attacks are considered difficult



```
return!0<t?<3?crypt(-79,-13,+crypt(-87,1-_,crypt(-86,0,a1)+a)):
1,t<_2crypt(t+1,_,a):3,crypt(-94,-27+t,a)&&t==22<13?
crypt(2,_,1,"%s %d %d\n"):9:16:t<0?t<-?2?crypt(,t,
"@n'+,#'/*{}/w#wcdnr/+,}{r/*de}+,/*{**+,/w{**+,/w#q#n+,#l+,/n(n+,/+n+,/#\
#q#n+/,+k#;+*,/'r : 'de'3,}{w+K w'K'+)e#;dq#'\ \
q#'+d'K#!/+k#;q#'+)eK#}w'r)eK#(n!/'#;#q#n''){}#jw''){}(n!)/+n';d)rw' i;#\
}{n!}/n{#': r{#w'r nc(n!)/#L+'K {rw' iK;{(n!)/w#q#n'wk mw' \
i#wk{K(n!)/w{#l#m#'+ i; :{n!}'/*{q#ld;r'}{nwb!/*de}'c \
; ;{n!-(}{rw' /+,)##*#}nc, '#mw' /+kde'+e);#rdq#w! nr' / ')}+(r!#('n' ')# \
}'+'#)##(!'!'):t<-50?_==a?putchar(31[a]):crypt(-65,_,a+1):crypt(#{a== '/' +t,_,a+1)
:0<t?crypt(2,2,"$s"):a== '/' || crypt(0,crypt(-61,_,a,
"!ek;dc i@k'(q)-(w)*n+r3#l,{};mwLoca-0m .vbvks,fxntdCeqhiry",a+1);
```



Why data-only attacks are **considered difficult**



How data-only attacks **can be**



Why data-only attacks are **considered difficult**

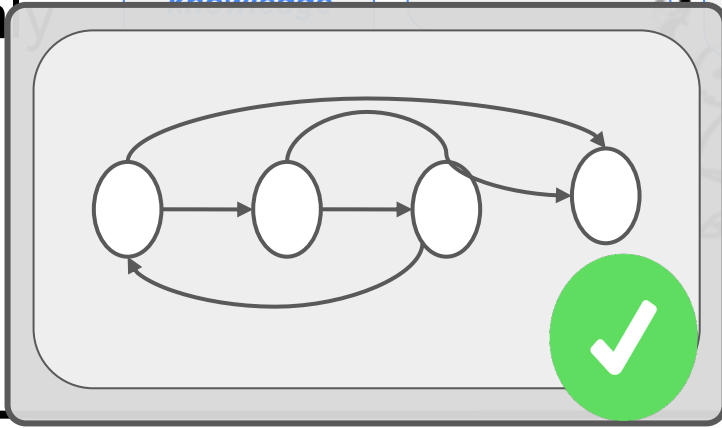
Application-specific knowledge

Heavyweight analyses

Complex dataflow constraints

Bypass defenses

Turing-complete gadget set



How data-only attacks **can be**

Victim executes its intended code



Application-specific knowledge

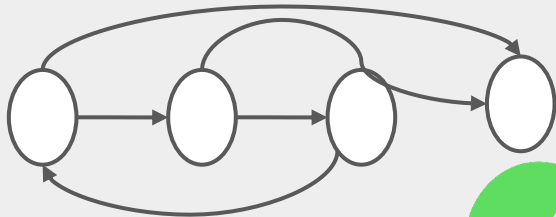
Heavyweight analyses

Complex dataflow constraints

Bypass defenses

Turing-complete gadget set

Why data-only attacks are considered difficult



`cgi_bin: "/usr/cgi-bin"`



`execve(path="/usr/cgi-bin...", ...);`

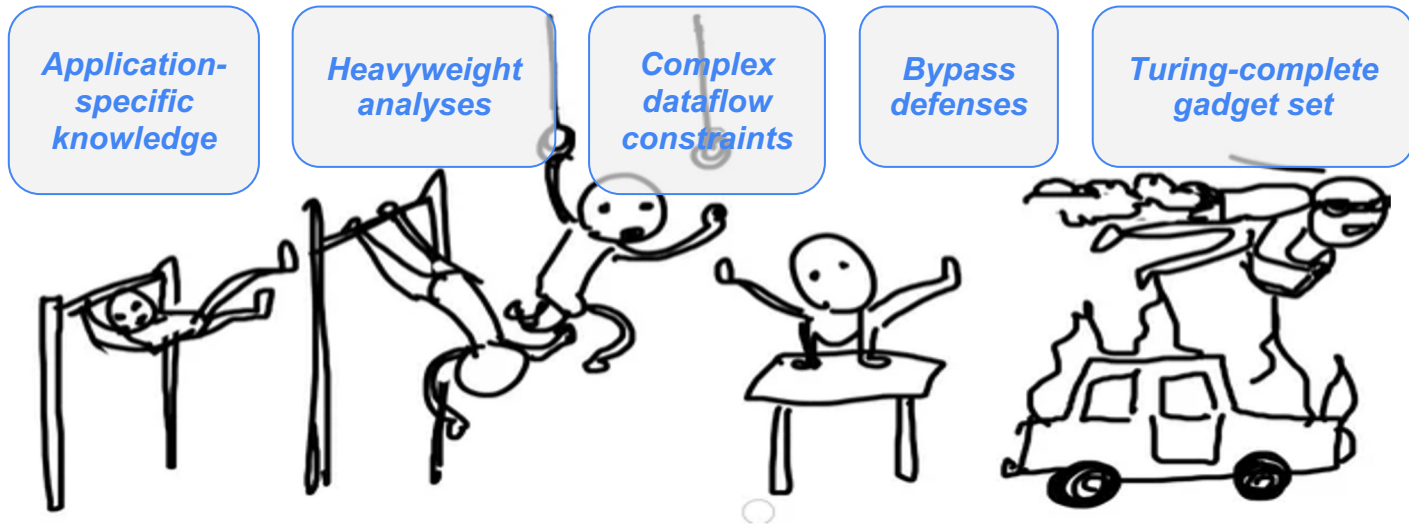
Victim executes its intended code

Victim passes attacker data to a syscall verbatim

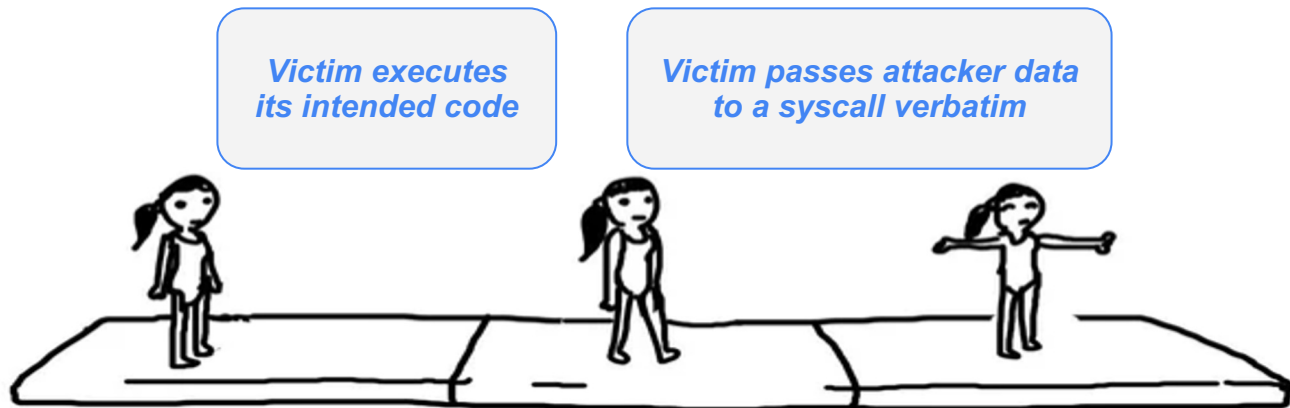
How data-only attacks can be



Why data-only attacks are **considered difficult**



How data-only attacks can be



Einstein: “As simple as possible, but not simpler”

Einstein: “As simple as possible, but not simpler”

① *Which data to overwrite?*

Einstein: “As simple as possible, but not simpler”

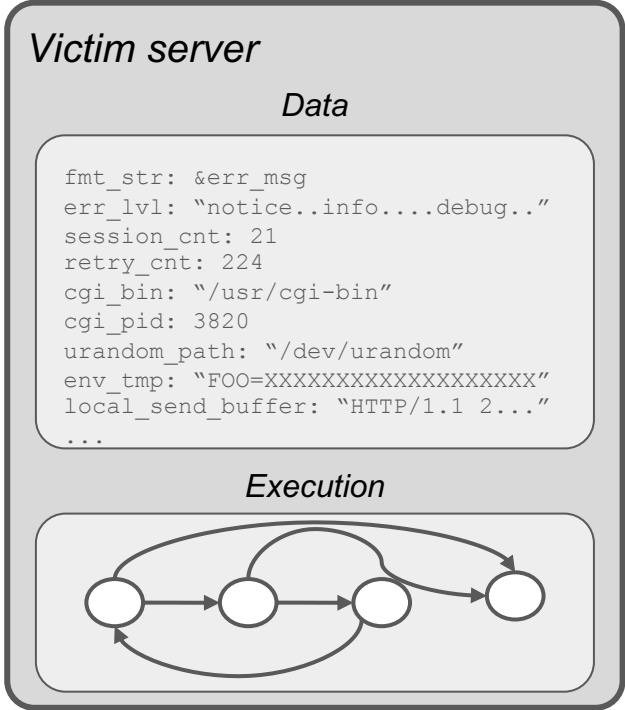
① *Which data to overwrite?*

② *What to overwrite it with?*

Einstein: “As simple as possible, but not simpler”

① *Which data to overwrite?*

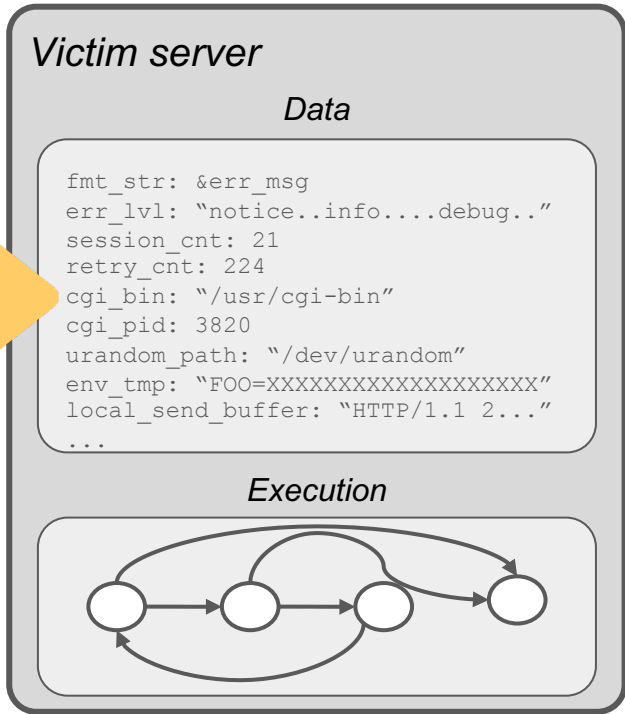
② *What to overwrite it with?*



Einstein: "As simple as possible, but not simpler"

① *Which data to overwrite?*

② *What to overwrite it with?*

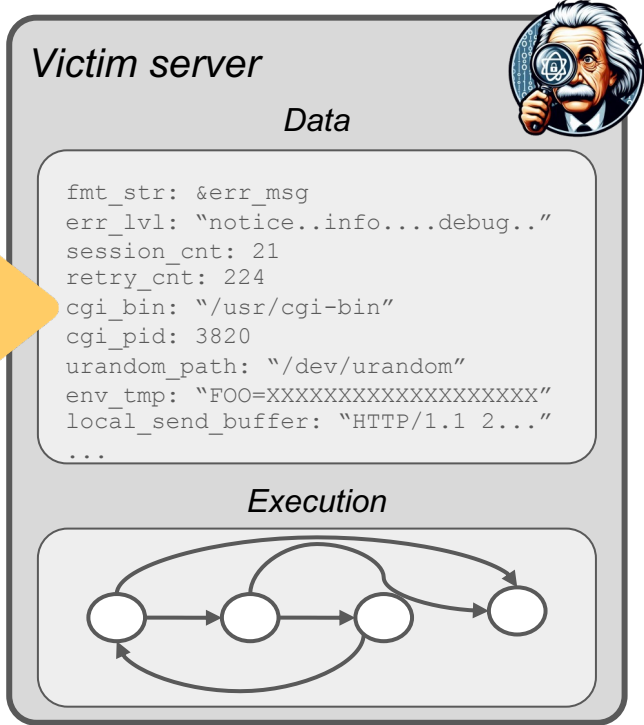


Einstein: "As simple as possible, but not simpler"

⊙ *Which data to overwrite?*

💡 *Use dynamic taint analysis*

⊙ *What to overwrite it with?*

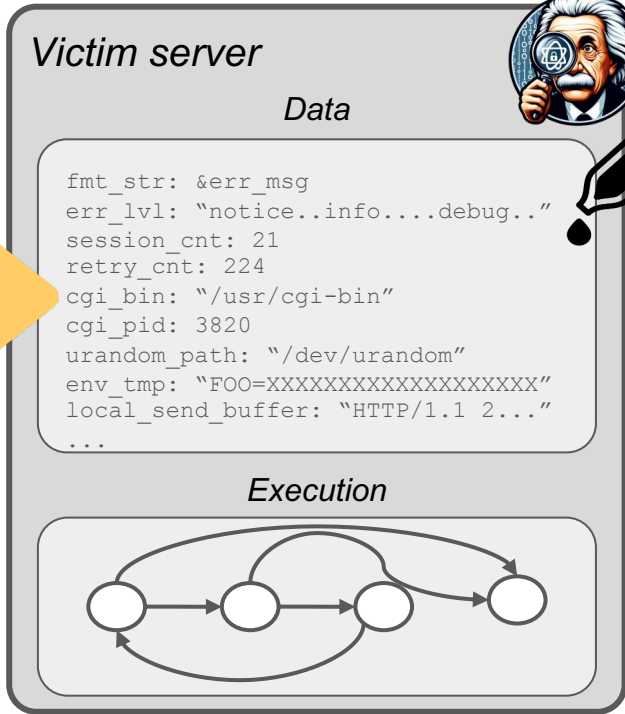


Einstein: "As simple as possible, but not simpler"

⓪ *Which data to overwrite?*

⓪ *What to overwrite it with?*

💡 *Use **dynamic taint analysis***

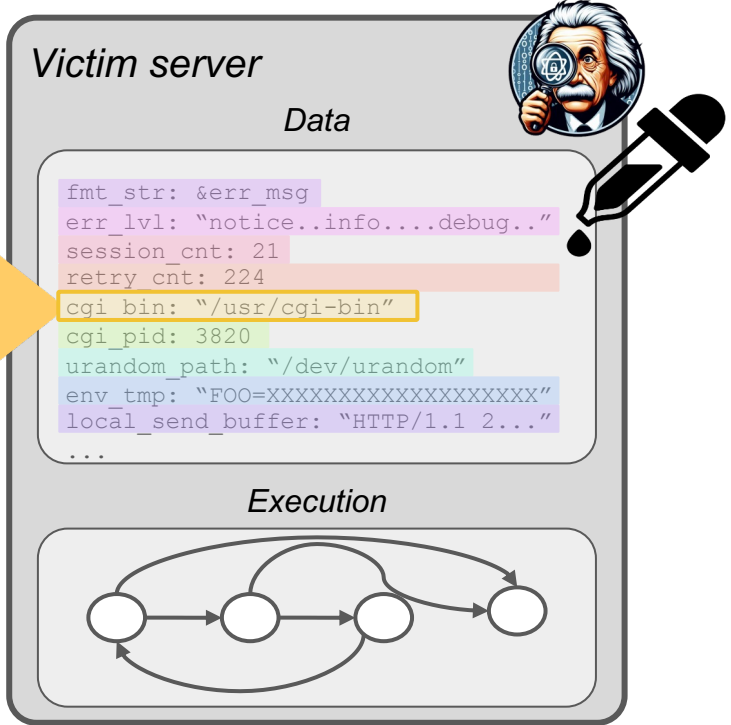


Einstein: "As simple as possible, but not simpler"

⊛ *Which data to overwrite?*

⊛ *What to overwrite it with?*

💡 - Use **dynamic taint analysis**

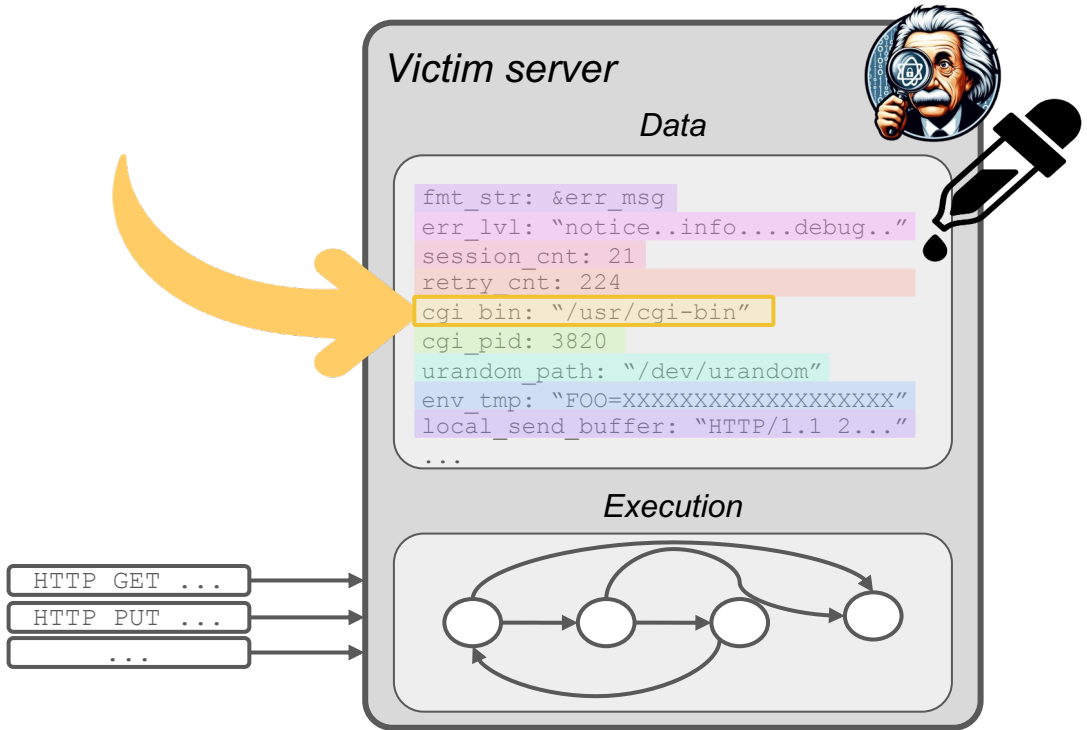


Einstein: "As simple as possible, but not simpler"

⊛ Which data to overwrite?

⊛ What to overwrite it with?

💡 Use **dynamic taint analysis**

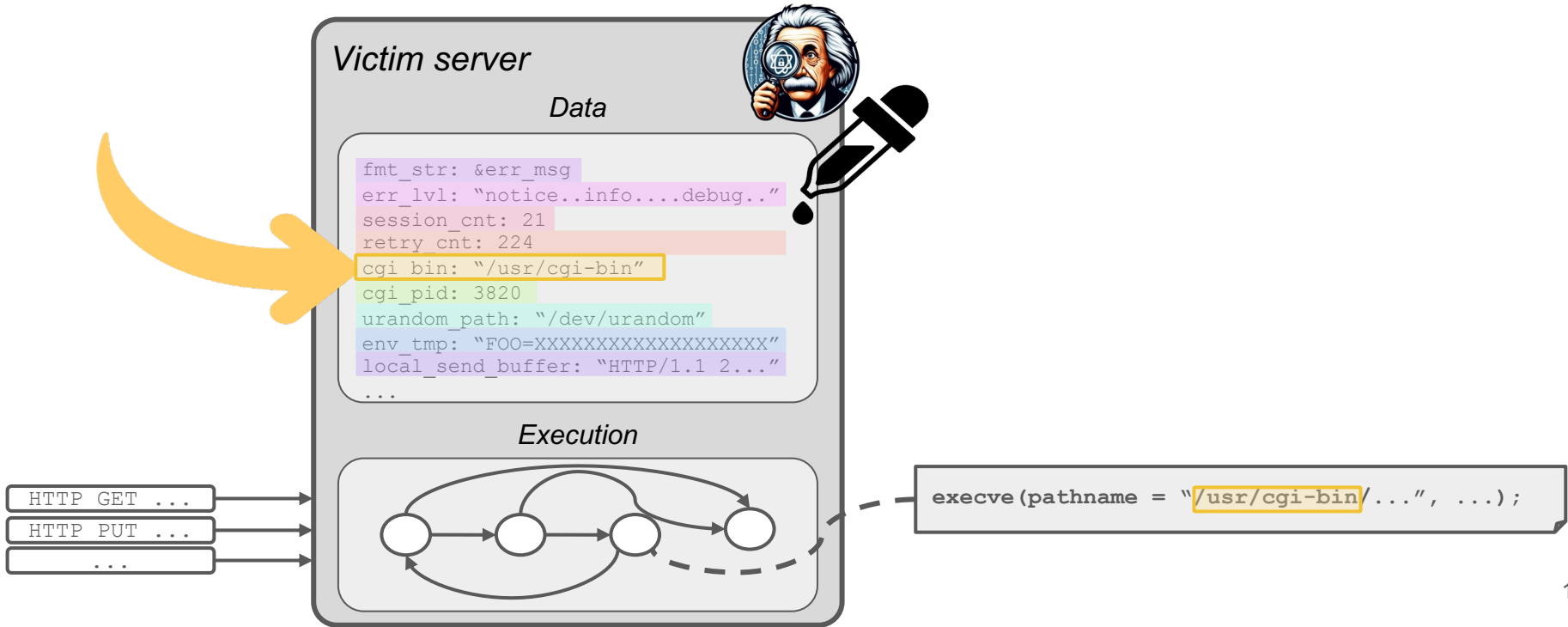


Einstein: "As simple as possible, but not simpler"

⊛ Which data to overwrite?

⊛ What to overwrite it with?

💡 Use **dynamic taint analysis**

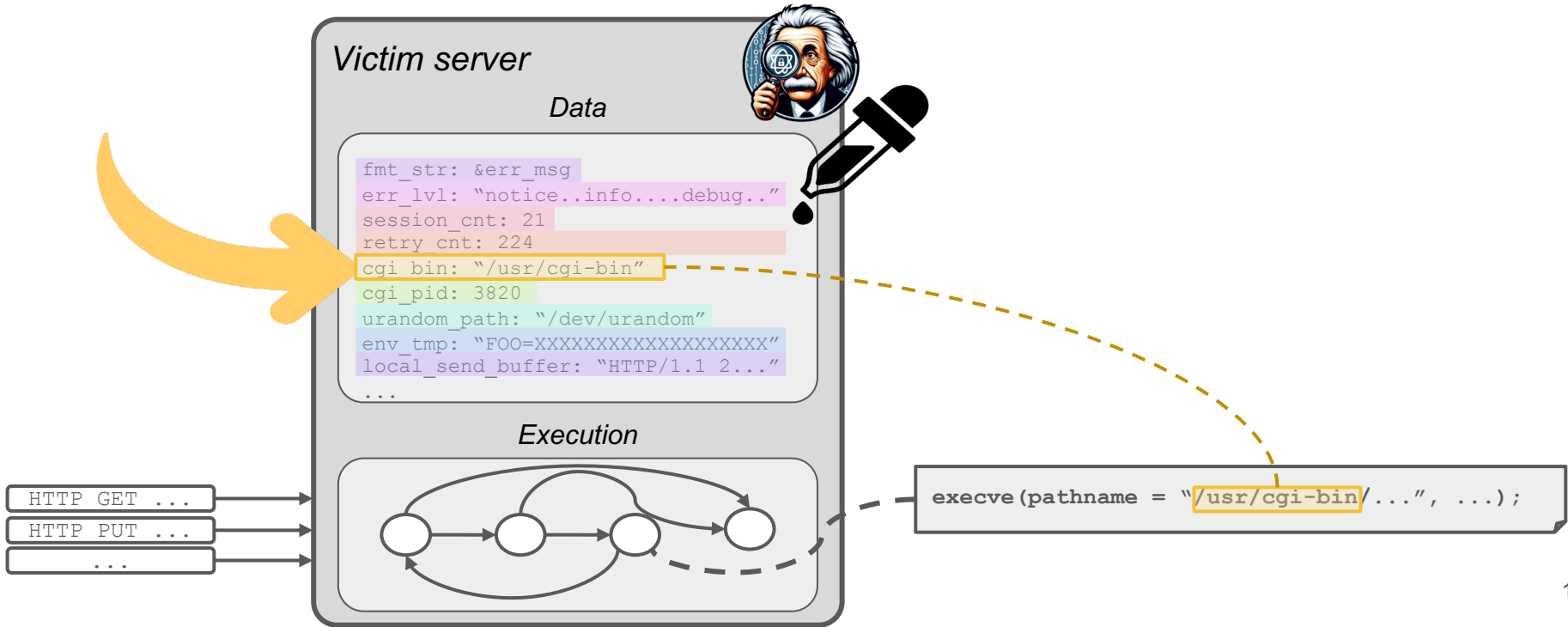


Einstein: "As simple as possible, but not simpler"

⊛ Which data to overwrite?

⊛ What to overwrite it with?

💡 Use **dynamic taint analysis**

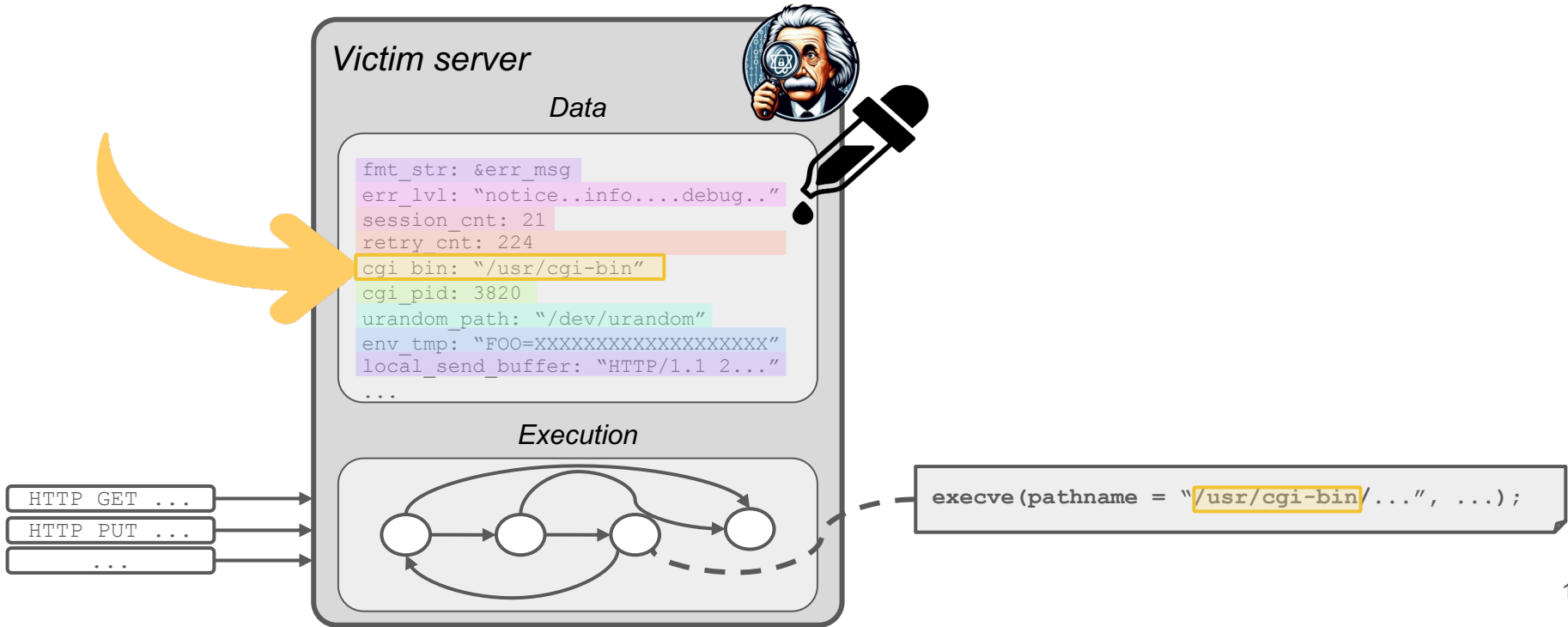


Einstein: "As simple as possible, but not simpler"

⊛ Which data to overwrite?

⊛ What to overwrite it with?

💡 Use **dynamic taint analysis**



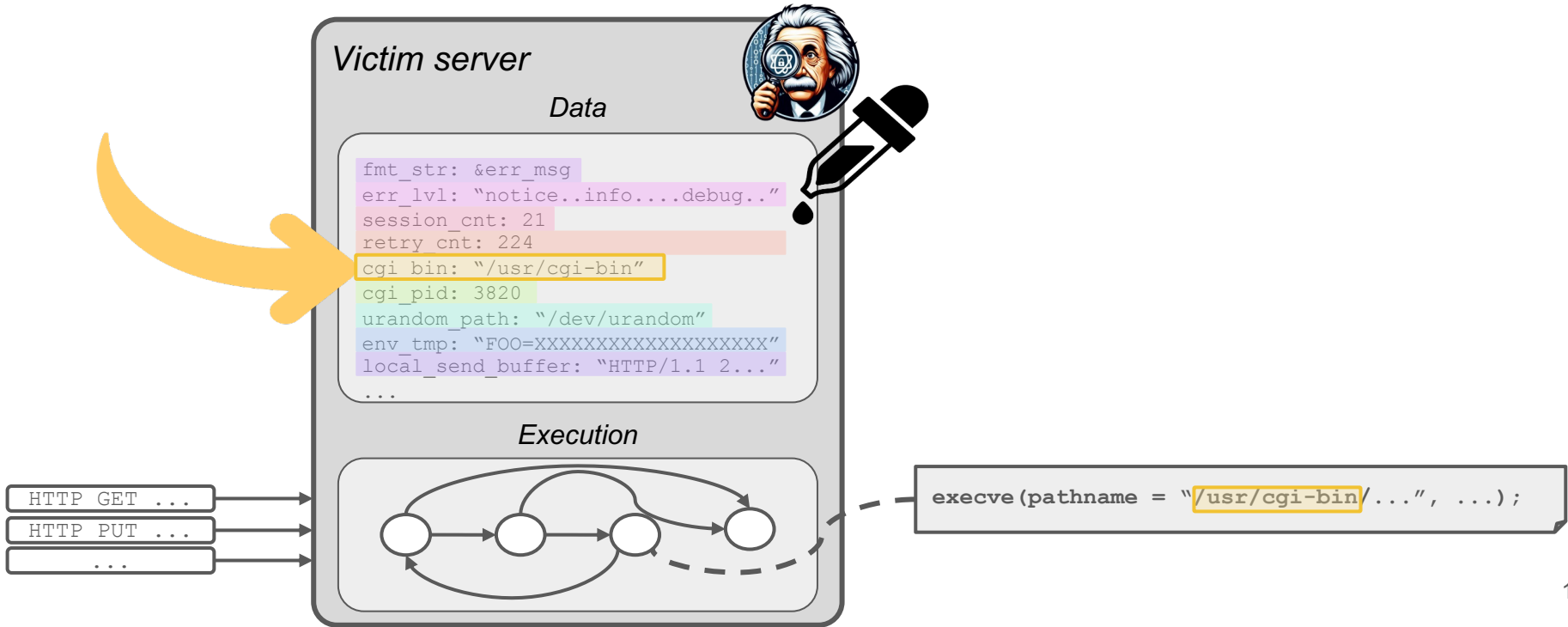
Einstein: "As simple as possible, but not simpler"

? Which data to overwrite?

-💡- Use **dynamic taint analysis**

? What to overwrite it with?

-💡- Exploit data copied **verbatim**



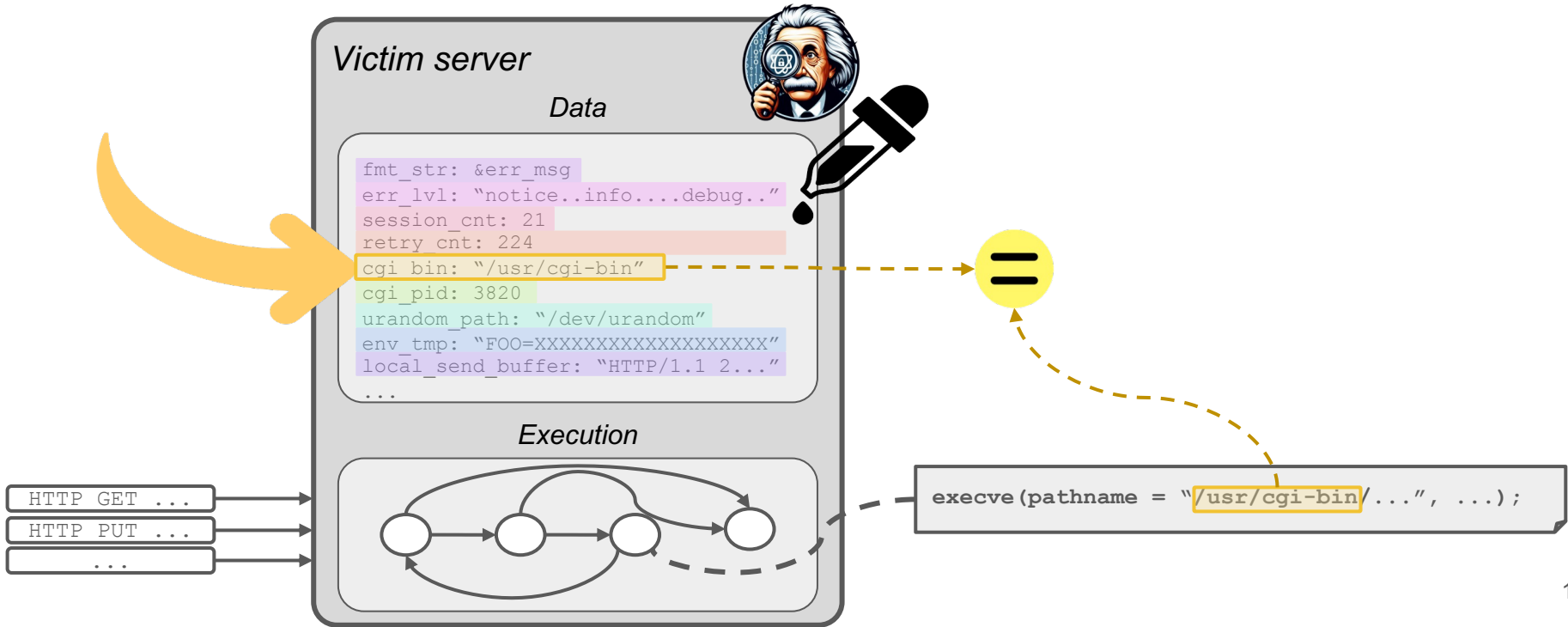
Einstein: "As simple as possible, but not simpler"

? Which data to overwrite?

-💡- Use **dynamic taint analysis**

? What to overwrite it with?

-💡- Exploit data copied **verbatim**



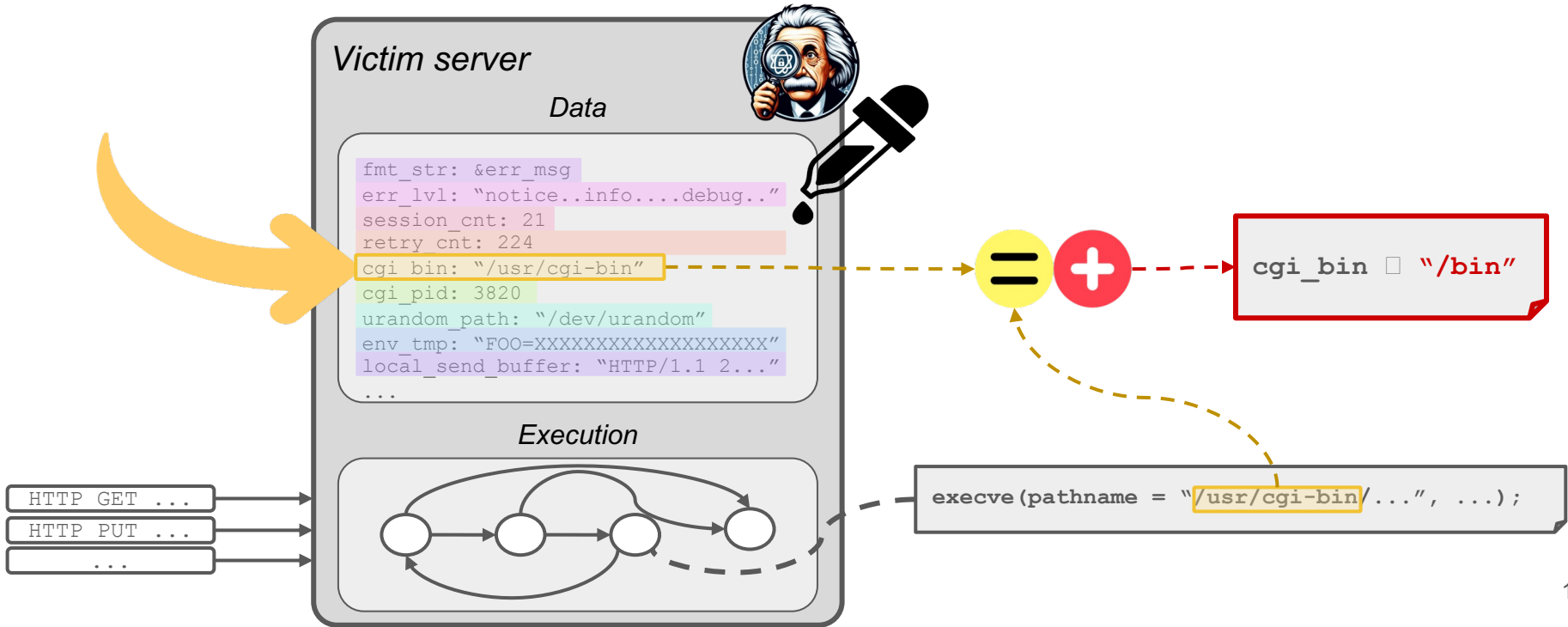
Einstein: "As simple as possible, but not simpler"

? Which data to overwrite?

-💡- Use **dynamic taint analysis**

? What to overwrite it with?

-💡- Exploit data copied **verbatim**



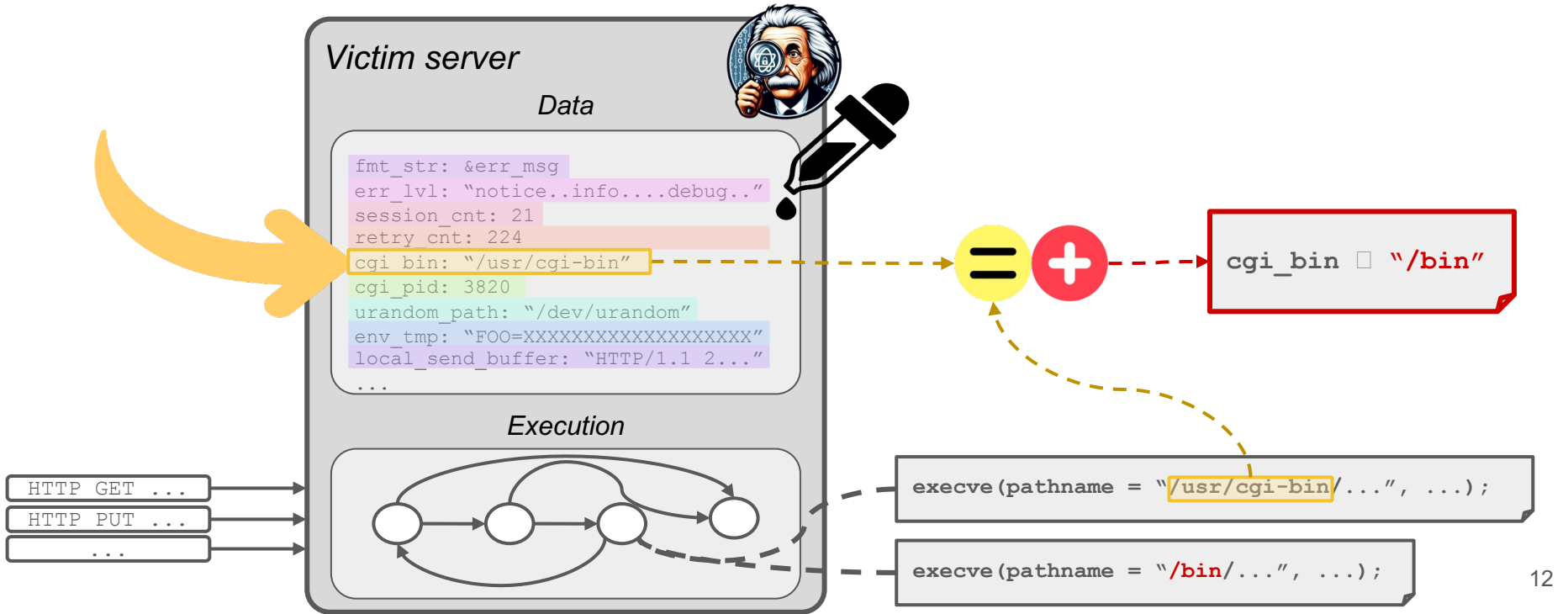
Einstein: "As simple as possible, but not simpler"

? Which data to overwrite?

-💡- Use **dynamic taint analysis**

? What to overwrite it with?

-💡- Exploit data copied **verbatim**



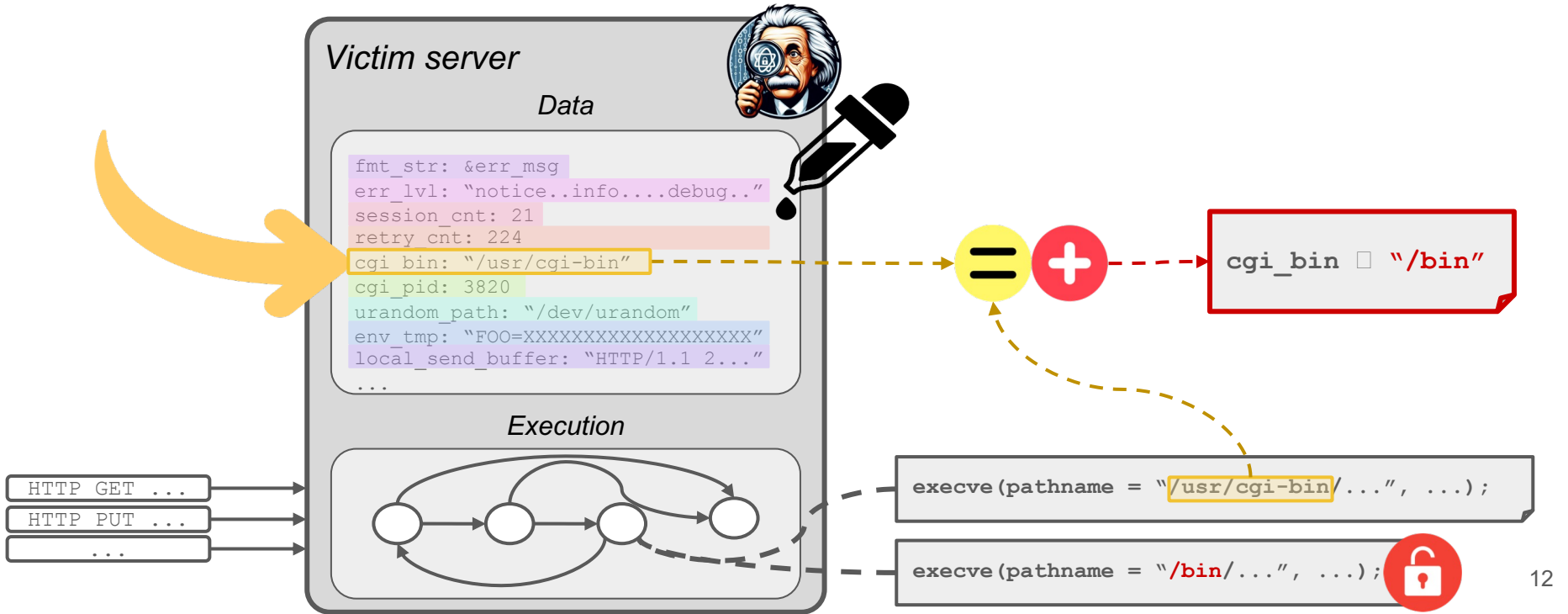
Einstein: "As simple as possible, but not simpler"

? Which data to overwrite?

-💡- Use **dynamic taint analysis**

? What to overwrite it with?

-💡- Exploit data copied **verbatim**



Evaluation: Setup

Evaluation: Setup

Target applications:

Evaluation: Setup

Target applications:



Evaluation: Setup

Target applications:



- CVE-2022-23943
- CVE-2022-41741
- CVE-2007-4727
- CVE-2023-36824
- CVE-2021-32027
- ...



Evaluation: Setup

Target applications:



- CVE-2022-23943
- CVE-2022-41741
- CVE-2007-4727
- CVE-2023-36824
- CVE-2021-32027
- ...



Workloads:

Evaluation: Setup

Target applications:



- CVE-2022-23943
- CVE-2022-41741
- CVE-2007-4727
- CVE-2023-36824
- CVE-2021-32027
- ...



Workloads: Each application's test suite

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
<code>execve</code>	11	7 (86%)	7 (86%)	7 (86%)			

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
<code>execve</code>	11	7 (86%)	7 (86%)	7 (86%)			

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
<code>execve</code>	11	7 (86%)	7 (86%)	7 (86%)			

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
execve	11	7 (86%) <i>pathname</i>	7 (86%) <i>argv</i>	7 (86%) <i>envp</i>			

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
<code>execve</code>	11	7 (86%) <i>pathname</i>	7 (86%) <i>argv</i>	7 (86%) <i>envp</i>			

After initialization,
programs typically copy
strings around *verbatim*.

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
<code>execve</code>	11	7 (86%) <i>pathname</i>	7 (86%) <i>argv</i>	7 (86%) <i>envp</i>			
<code>mmap</code>	787	55 (5%)	441 (8%)	55	16 (100%)	17	73
<code>mprotect</code>	144	97 (68%)	98 (27%)	1			
<code>mremap</code>	11	11	7	11	-		
<code>pwrite64</code>	1270	1217 (3%)	741 (47%)	399 (19%)	506 (36%)		
<code>pwritev</code>	10	10 (100%)	10 (10%)	-	10 (20%)		
<code>sendfile</code>	1	-	-	1	1		
<code>sendmmsg</code>	2	2	2	-	-		
<code>sendmsg</code>	12	5 (100%)	3 (100%)	-			
<code>sendto</code>	431	389 (7%)	410 (38%)	408 (6%)	-	-	-
<code>write</code>	3083	768 (74%)	2877 (91%)	1265 (10%)			
<code>writev</code>	754	244 (18%)	742 (76%)	-			

*After initialization, programs typically copy **strings** around **verbatim**.*

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
execve	11	7 (86%) <i>pathname</i>	7 (86%) <i>argv</i>	7 (86%) <i>envp</i>			
mmap	787	55 (5%)	441 (8%)	55	16 (100%)	17	73
mprotect	144	97 (68%)	98 (27%)	1			
mremap	11	11	7	11	-		
pwrite64	1270	1217 (3%)	741 (47%)	399 (19%)	506 (36%)		
pwritev	10	10 (100%)	10 (10%)	-	10 (20%)		
sendfile	1	-	-	1	1		
sendmmsg	2	2	2	-	-		
sendmsg	12	5 (100%)	3 (100%)	-			
sendto	431	389 (7%)	410 (38%)	408 (6%)	-	-	-
write	3083	768 (74%)	2877 (91%)	1265 (10%)			
writev	754	244 (18%)	742 (76%)	-			

*After initialization, programs typically copy **strings** around **verbatim**.*

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
execve	11	7 (86%) <i>pathname</i>	7 (86%) <i>argv</i>	7 (86%) <i>envp</i>			
mmap	787	55 (5%)	441 (8%)	55	16 (100%)	17	73
mprotect	144	97 (68%)	98 (27%)	1			
mremap	11	11	7	11	-		
pwrite64	1270	1217 (3%)	741 (47%)	399 (19%)	506 (36%)		
pwritev	10	10 (100%)	10 (10%)	-	10 (20%)		
sendfile	1	-	-	1	1		
sendmmsg	2	2	2	-	-		
sendmsg	12	5 (100%)	3 (100%)	-			
sendto	431	389 (7%)	410 (38%)	408 (6%)	-		
write	3083	768 (74%)	2877 (91%)	1265 (10%)			
writev	754	244 (18%)	742 (76%)	-			

*After initialization, programs typically copy **strings** around **verbatim**.*

*Many types of syscalls offer many **primitives** to an attacker.*

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
execve	11	7 (86%) <i>pathname</i>	7 (86%) <i>argv</i>	7 (86%) <i>envp</i>			
mmap	787	55 (5%)	441 (8%)	55	16 (100%)	17	73
mprotect	144	97 (68%)	98 (27%)	1			
mremap	11	11	7	11	-		
pwrite64	1270	1217 (3%)	741 (47%)	399 (19%)	506 (36%)		
pwritev	10	10 (100%)	10 (10%)	-	10 (20%)		
sendfile	1	-	-	1	1		
sendmmsg	2	2	2	-	-		
sendmsg	12	5 (100%)	3 (100%)	-			
sendto	431	389 (7%)	410 (38%)	408 (6%)	-		
write	3083	768 (74%)	2877 (91%)	1265 (10%)			
writev	754	244 (18%)	742 (76%)	-			

*After initialization, programs typically copy **strings** around **verbatim**.*

*Many types of syscalls offer many **primitives** to an attacker.*

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
execve	11	7 (86%) <i>pathname</i>	7 (86%) <i>argv</i>	7 (86%) <i>envp</i>			
mmap	787	55 (5%)	441 (8%)	55	16 (100%)	17	73
mprotect	144	97 (68%)	98 (27%)	1			
mremap	11	11	7	11	-		
pwrite64	1270	1217 (3%)	741 (47%)	399 (19%)	506 (36%)		
pwritev	10	10 (100%)	10 (10%)	-	10 (20%)		
sendfile	1	-	-	1	1		
sendmmsg	2	2	2	-	-		
sendmsg	12	5 (100%)	3 (100%)	-			
sendto	431	389 (7%)	410 (38%)	408 (6%)	-		
write	3083	768 (74%)	2877 (91%)	1265 (10%)			
writev	754	244 (18%)	742 (76%)	-			

*After initialization, programs typically copy **strings** around **verbatim**.*

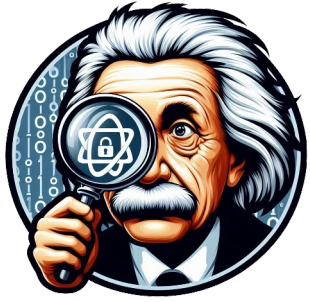
*Many types of syscalls offer many **primitives** to an attacker.*

Evaluation

Syscall	Total covered	Total attacker-tainted args. (% that are copied from attacker-controllable data <i>verbatim</i>)					
		Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5	Arg. 6
execve	11	7 (86%) <i>pathname</i>	7 (86%) <i>argv</i>	7 (86%) <i>envp</i>			
mmap	787	55 (5%)	441 (8%)	55	16 (100%)	17	73
mprotect	144	97 (68%)	98 (27%)	1			
mremap	11	11	7	11	-		
pwrite64	1270	1217 (3%)	741 (47%)	399 (19%)	506 (36%)		
pwritev	10	10 (100%)	10 (10%)	-	10 (20%)		
sendfile	1	-	-	1	1		
sendmmsg	2	2	2	-	-		
sendmsg	12	5 (100%)	3 (100%)	-			
sendto	431	389 (7%)	410 (38%)	408 (6%)	-		
write	3083	768 (74%)	2877 (91%)	1265 (10%)			
writerv	754	244 (18%)	742 (76%)	-			

*After initialization, programs typically copy **strings** around **verbatim**.*

*Many types of syscalls offer many **primitives** to an attacker.*







?

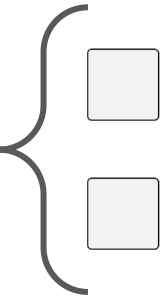








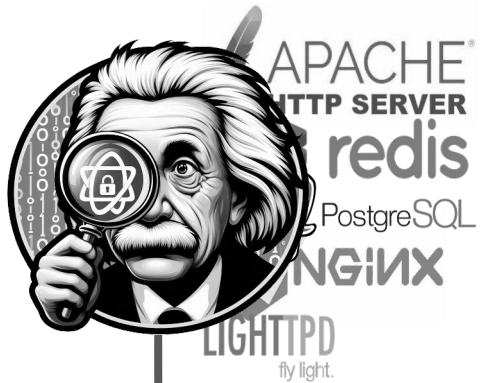
?





Comprehensive



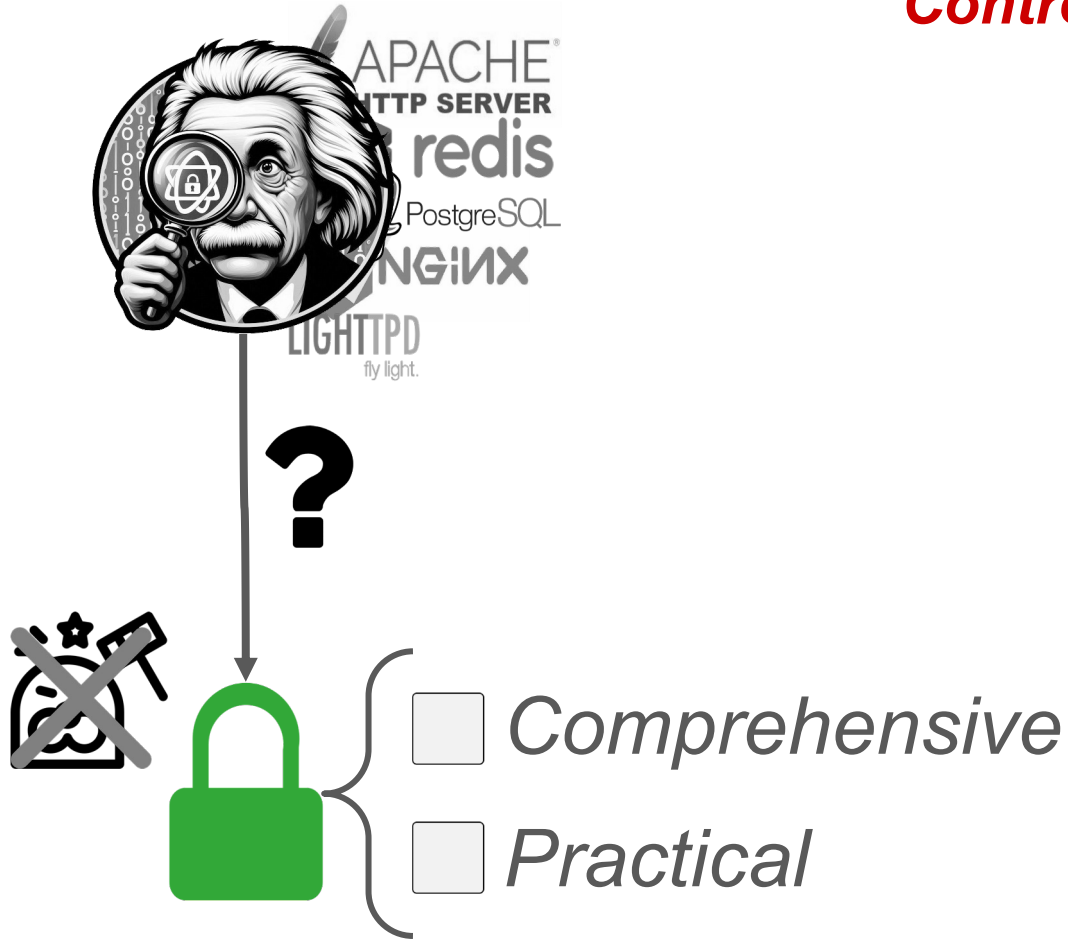


Comprehensive



Practical

Control-flow hijacking attacks



Control-flow hijacking attacks



Code pointer



Comprehensive



Practical

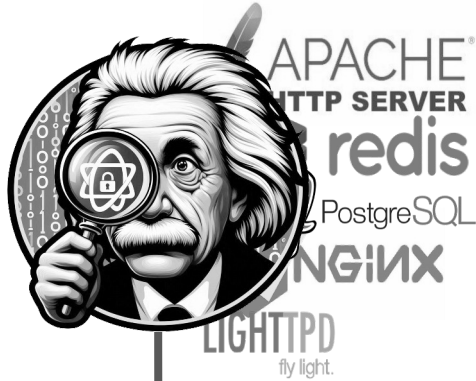
Control-flow hijacking attacks



Code pointer



Indirect branch



Comprehensive



Practical



Comprehensive

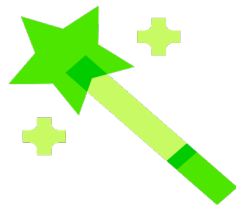
Practical

Control-flow hijacking attacks



Code pointer

Indirect branch



Control-flow hijacking attacks



Code pointer



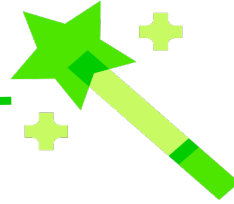
Indirect branch



Comprehensive



Practical





Comprehensive



Practical



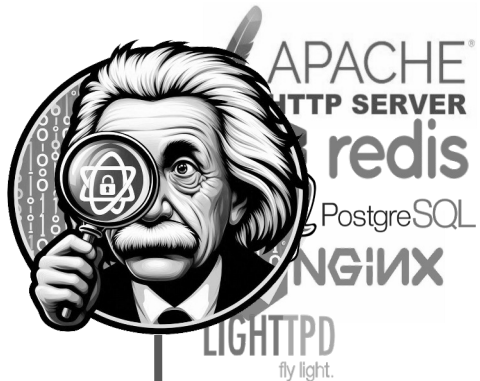
Comprehensive



Practical

Data-only attacks





Comprehensive

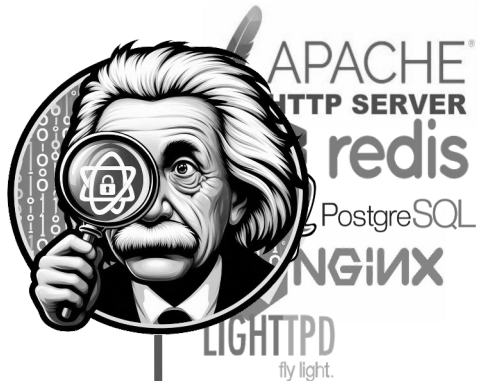


Practical

Data-only attacks



Any string
Any number
Any boolean
...



Comprehensive



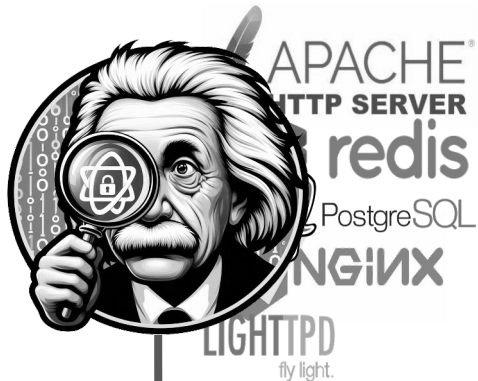
Practical

Data-only attacks



Any string
Any number
Any boolean
...

Any syscall
Any store
Any cond. branch
...



Comprehensive

Practical

Data-only attacks



Any string
Any number
Any boolean
...

Any syscall
Any store
Any cond. branch
...





Comprehensive



Practical

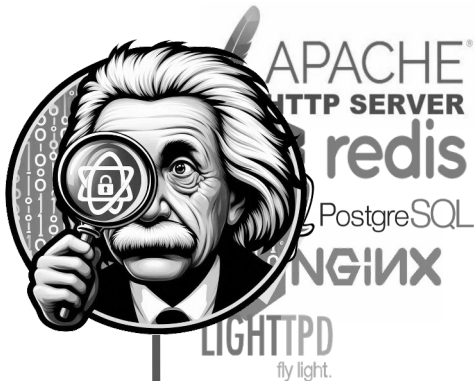
Data-only attacks



Any string
Any number
Any boolean
...

Any syscall
Any store
Any cond. branch
...





Comprehensive



Practical



Any string
Any number
Any boolean
...

Any syscall
Any store
Any cond. branch
...

Data-only attacks



Comprehensive



Practical

Data-only attacks



Any string
Any number
Any boolean
...

Any syscall
Any store
Any cond. branch
...





?



Comprehensive



Practical

Data-only attacks



Any string
Any number
Any boolean
...

Any syscall
Any store
Any cond. branch
...





*We present Einstein, a **data-only attack** generator.*



*We present Einstein, a **data-only attack** generator.*



*It builds **hundreds of exploits** against popular web servers.*



*We present Einstein, a **data-only attack** generator.*



*It builds **hundreds of exploits** against popular web servers.*



*We call upon researchers and vendors to **rethink mitigation strategies.***