Keeping customer data safe in EC2 – a deep dive

Martin Pohlack

Amazon Web Services

Rated PG13:
mild use of assembly

Bio ...

- Principal Engineer with Amazon Web Services
- I like to play with
  - Low-level stuff
  - Synchronization, hardware transactional memory
  - Virtualization
  - Real-time systems, micro-kernel systems
  - Reactive security

Keeping customer data safe

- Security is tenet #1 in AWS
- Focus: issues in Xen virtualization stack
- Example: a Xen security advisory

Xen Security Advisory CVE-2015-2151 / ***XSA-123***

Hypervisor memory corruption due to x86 emulator flaw

*** EMBARGOED UNTIL ***2015-03-10 12:00 UTC*** ***

ISSUE DESCRIPTION

Instructions with register operands ignore eventual segment Overrides encoded for them. Due to an ***insufficiently conditional assignment*** such a bogus segment override can, however, ***corrupt a pointer*** [...]

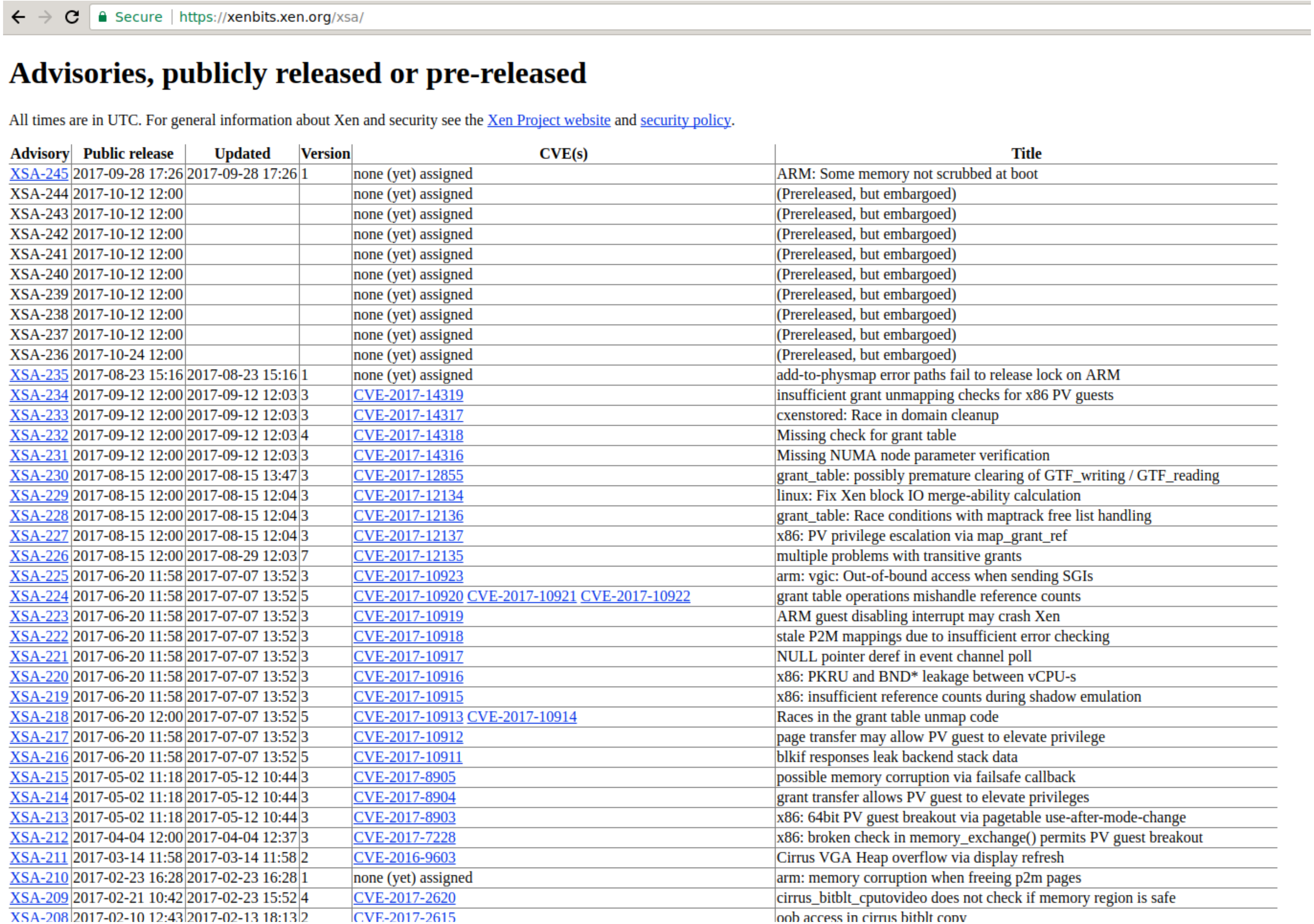
IMPACT

A malicious guest might be able to ***read sensitive data*** relating to other guests, or to cause ***denial of service*** on the host. Arbitrary code execution, and therefore ***privilege escalation, cannot be excluded***.

VULNERABLE SYSTEMS: ***Xen 3.2.x and later are vulnerable.***

MITIGATION: There is ***no mitigation available*** for this issue.

RESOLUTION: xsa123-4.3-4.2.patch Xen 4.3.x, Xen 4.2.x



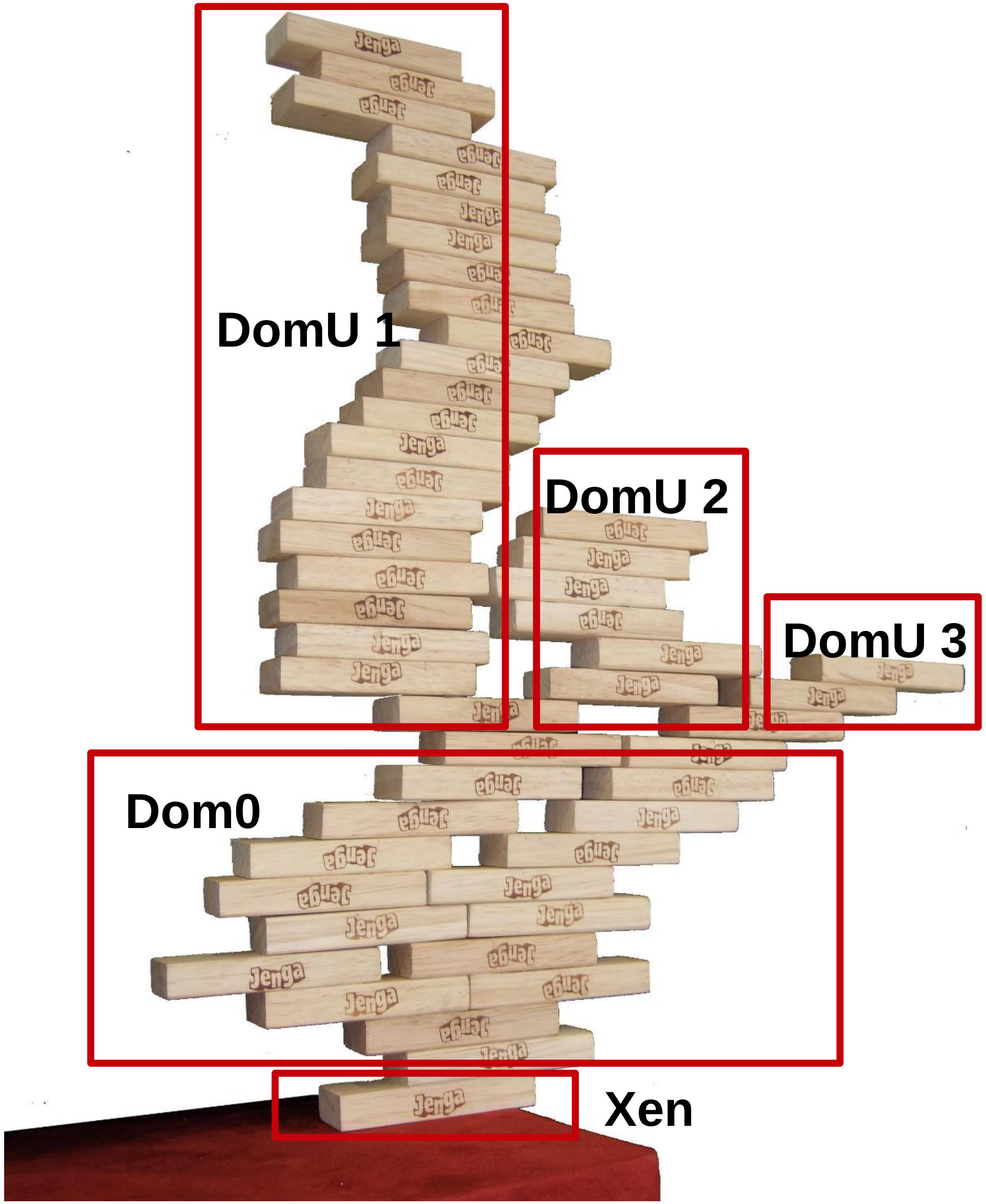
Advisories, publicly released or pre-released

All times are in UTC. For general information about Xen and security see the Xen Project website and security policy.

Advisory	Public release	Updated	Version	CVE(s)	Title
XSA-245	2017-09-28 17:26	2017-09-28 17:26	1	none (yet) assigned	ARM: Some memory not scrubbed at boot
XSA-244	2017-10-12 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-243	2017-10-12 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-242	2017-10-12 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-241	2017-10-12 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-240	2017-10-12 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-239	2017-10-12 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-238	2017-10-12 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-237	2017-10-12 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-236	2017-10-24 12:00			none (yet) assigned	(Prereleased, but embargoed)
XSA-235	2017-08-23 15:16	2017-08-23 15:16	1	none (yet) assigned	add-to-physmap error paths fail to release lock on ARM
XSA-234	2017-09-12 12:00	2017-09-12 12:03	3	CVE-2017-14319	insufficient grant unmapping checks for x86 PV guests
XSA-233	2017-09-12 12:00	2017-09-12 12:03	3	CVE-2017-14317	cxenstored: Race in domain cleanup
XSA-232	2017-09-12 12:00	2017-09-12 12:03	4	CVE-2017-14318	Missing check for grant table
XSA-231	2017-09-12 12:00	2017-09-12 12:03	3	CVE-2017-14316	Missing NUMA node parameter verification
XSA-230	2017-08-15 12:00	2017-08-15 13:47	3	CVE-2017-12855	grant_table: possibly premature clearing of GTF_writing / GTF_reading
XSA-229	2017-08-15 12:00	2017-08-15 12:04	3	CVE-2017-12134	linux: Fix Xen block IO merge-ability calculation
XSA-228	2017-08-15 12:00	2017-08-15 12:04	3	CVE-2017-12136	grant_table: Race conditions with maptrack free list handling
XSA-227	2017-08-15 12:00	2017-08-15 12:04	3	CVE-2017-12137	x86: PV privilege escalation via map_grant_ref
XSA-226	2017-08-15 12:00	2017-08-29 12:03	7	CVE-2017-12135	multiple problems with transitive grants
XSA-225	2017-06-20 11:58	2017-07-07 13:52	3	CVE-2017-10923	arm: vgic: Out-of-bound access when sending SGIs
XSA-224	2017-06-20 11:58	2017-07-07 13:52	5	CVE-2017-10920 CVE-2017-10921 CVE-2017-10922	grant table operations mishandle reference counts
XSA-223	2017-06-20 11:58	2017-07-07 13:52	3	CVE-2017-10919	ARM guest disabling interrupt may crash Xen
XSA-222	2017-06-20 11:58	2017-07-07 13:52	3	CVE-2017-10918	stale P2M mappings due to insufficient error checking
XSA-221	2017-06-20 11:58	2017-07-07 13:52	3	CVE-2017-10917	NULL pointer deref in event channel poll
XSA-220	2017-06-20 11:58	2017-07-07 13:52	3	CVE-2017-10916	x86: PKRU and BND* leakage between vCPU-s
XSA-219	2017-06-20 11:58	2017-07-07 13:52	3	CVE-2017-10915	x86: insufficient reference counts during shadow emulation
XSA-218	2017-06-20 12:00	2017-07-07 13:52	5	CVE-2017-10913 CVE-2017-10914	Races in the grant table unmap code
XSA-217	2017-06-20 11:58	2017-07-07 13:52	3	CVE-2017-10912	page transfer may allow PV guest to elevate privilege
XSA-216	2017-06-20 11:58	2017-07-07 13:52	5	CVE-2017-10911	blkif responses leak backend stack data
XSA-215	2017-05-02 11:18	2017-05-12 10:44	3	CVE-2017-8905	possible memory corruption via failsafe callback
XSA-214	2017-05-02 11:18	2017-05-12 10:44	3	CVE-2017-8904	grant transfer allows PV guest to elevate privileges
XSA-213	2017-05-02 11:18	2017-05-12 10:44	3	CVE-2017-8903	x86: 64bit PV guest breakout via pagetable use-after-mode-change
XSA-212	2017-04-04 12:00	2017-04-04 12:37	3	CVE-2017-7228	x86: broken check in memory_exchange() permits PV guest breakout
XSA-211	2017-03-14 11:58	2017-03-14 11:58	2	CVE-2016-9603	Cirrus VGA Heap overflow via display refresh
XSA-210	2017-02-23 16:28	2017-02-23 16:28	1	none (yet) assigned	arm: memory corruption when freeing p2m pages
XSA-209	2017-02-21 10:42	2017-02-23 15:52	4	CVE-2017-2620	cirrus_bitblt_cputovideo does not check if memory region is safe
XSA-208	2017-02-10 12:43	2017-02-13 18:13	2	CVE-2017-2615	oob access in cirrus bitblt copy

Components

- Xen virtualization stack
  - **Xen hypervisor**
  - QEMU
  - Dom0 Linux kernel
  - ...


DomU 1
DomU 2
DomU 3
Dom0
Xen

Security response options

- Vendor-specific options
- Configuration changes
- Patch and reboot
- Live migration
- Hot patching

Security & availability: How?

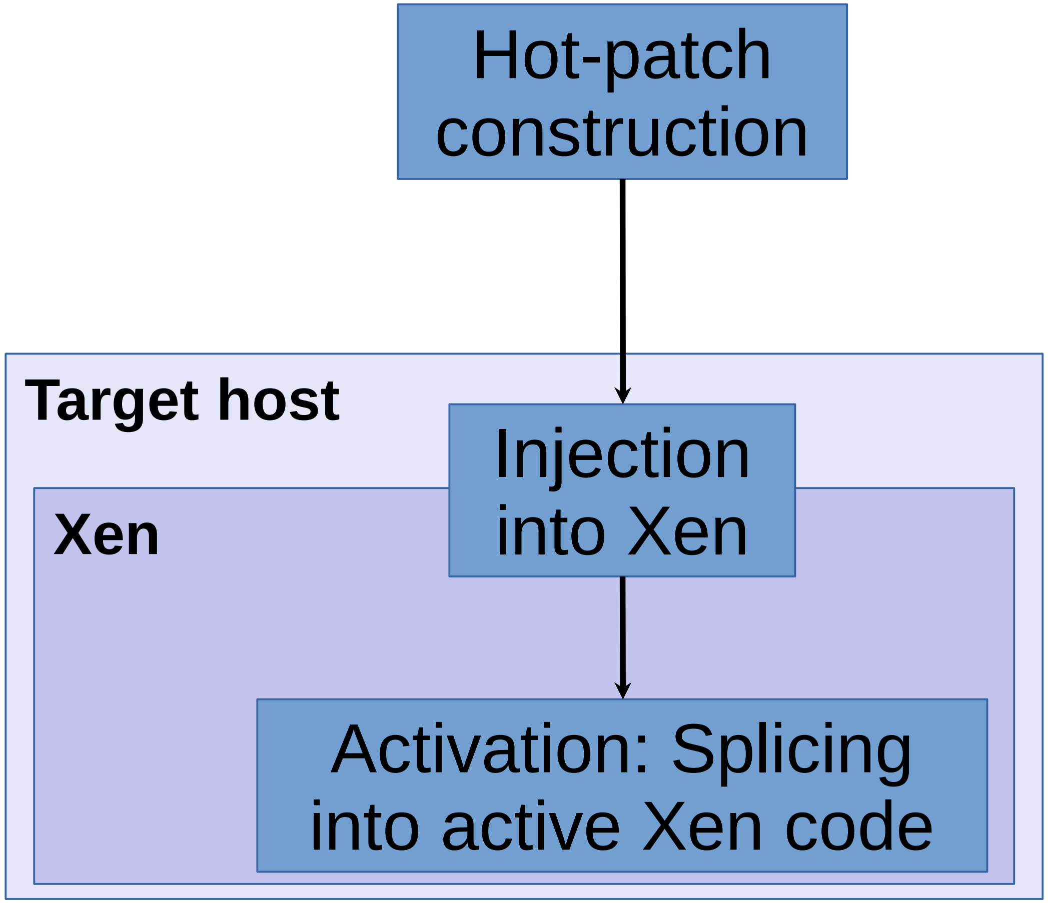
***Hot patching** [...] is the application of patches without shutting down and restarting the system **[...]**. This addresses problems related to **unavailability** of services **[...]**.*

https://en.wikipedia.org/wiki/Patch_(computing)#Hot_patching

Already solved?

- R. Wojtczuk: *Subverting the Xen hypervisor. Black Hat USA '08*
- J. Arnold, M. F. Kaashoek: *Ksplice: Automatic Rebootless Kernel Updates. EuroSys '09*
- kPatch (Redhat) +
  kgraft (SUSE) -> Linux livepatch (2014)
- Xen 4.7: Xen live patch, experimental (2016)
  Xen 4.9: supported on x86 (2017)

Building blocks



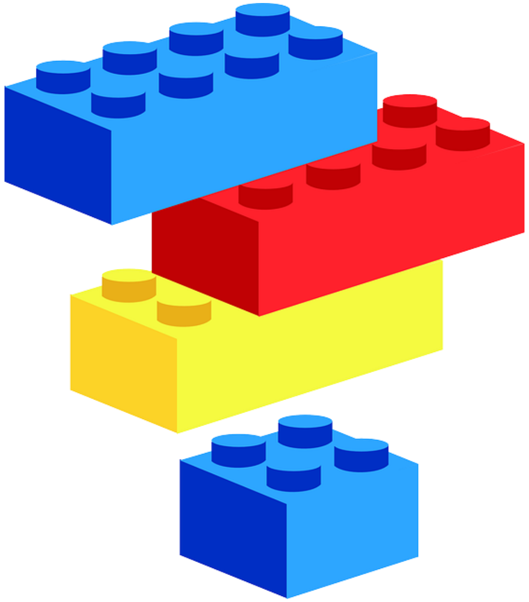
Hot-patch construction

Target host

Xen

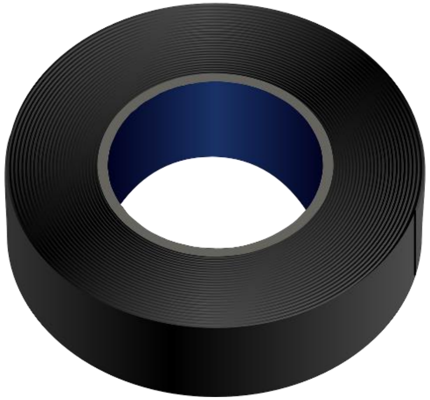
Injection into Xen

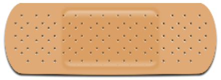
Activation: Splicing into active Xen code



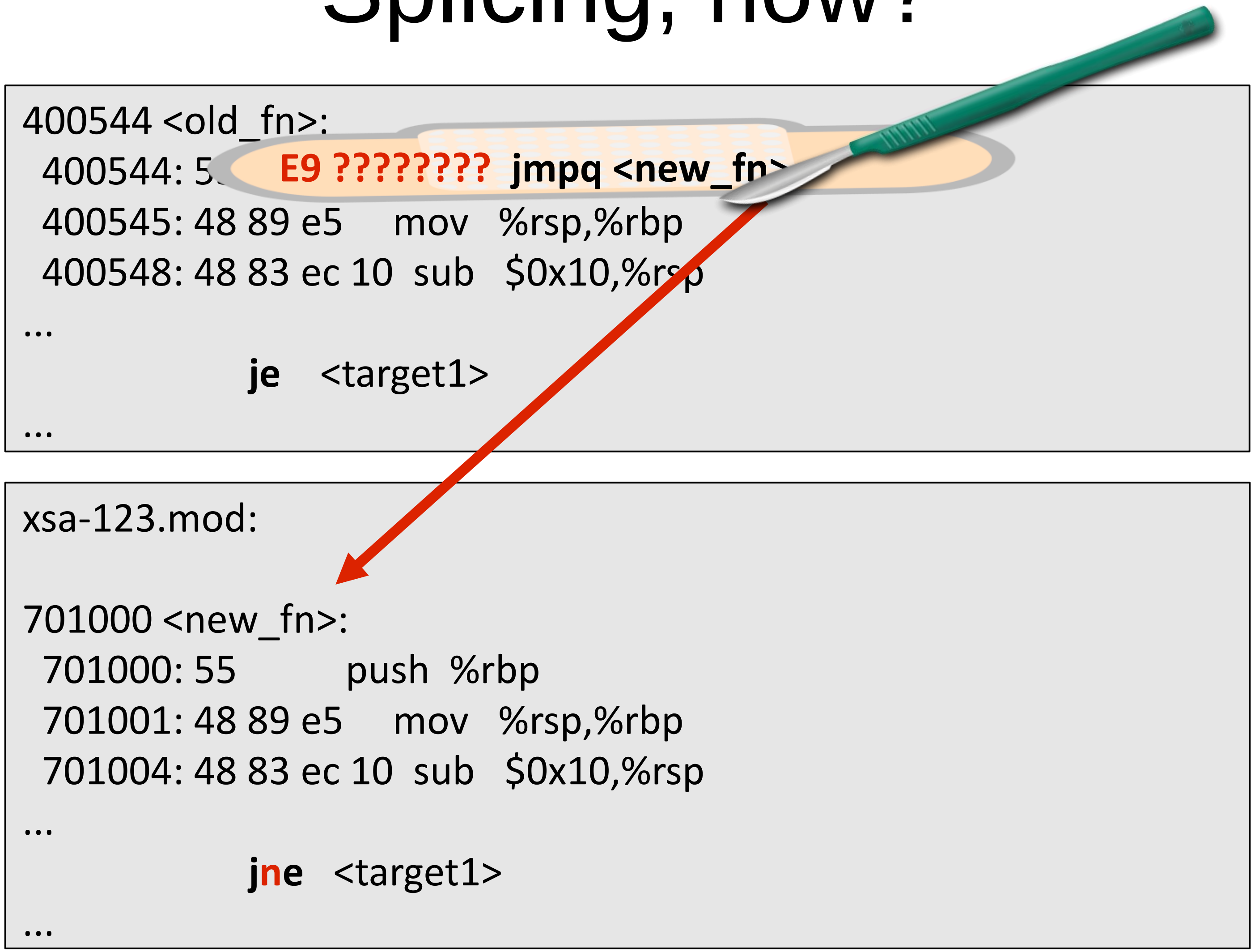
Xen under the hood ...

Splicing, what?





Splicing, how?



```
400544 <old_fn>:
  400544: 5  E9 ????????  jmpq <new_fn>
  400545: 48 89 e5     mov   %rsp,%rbp
  400548: 48 83 ec 10  sub   $0x10,%rsp
...
               je    <target1>
...
```

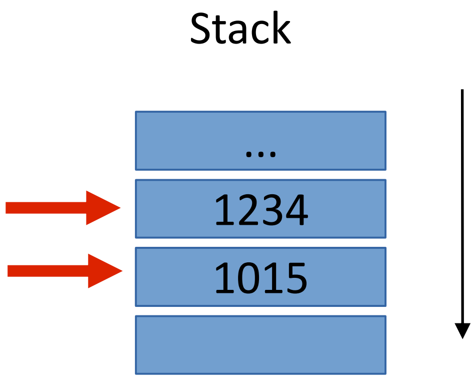
```
xsa-123.mod:

701000 <new_fn>:
  701000: 55           push  %rbp
  701001: 48 89 e5     mov   %rsp,%rbp
  701004: 48 83 ec 10  sub   $0x10,%rsp
...
               jne   <target1>
...
```

Splicing, when?

- Patch targets quiet
- Atomically

CPU stacks and function calls

```
1000 <f1>:
  ...
  1010:  call 2000 <f2>
  1015:  mov  ...
  ...
2000 <f2>:
  2000:  ...
  ...
  2100:  ret
```

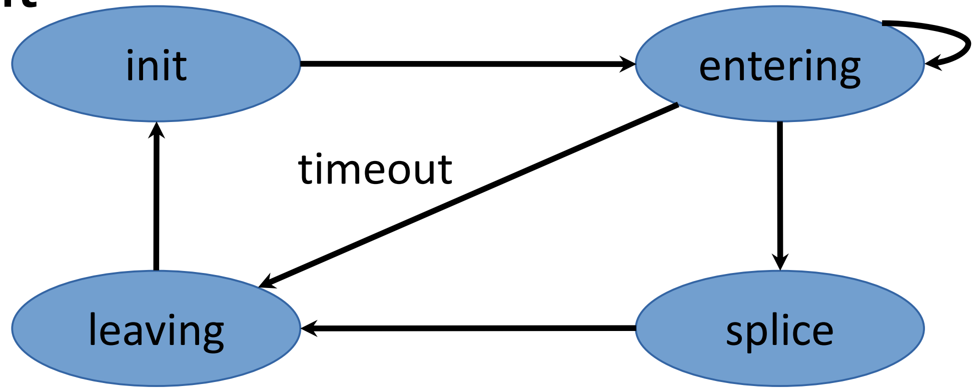


Stack

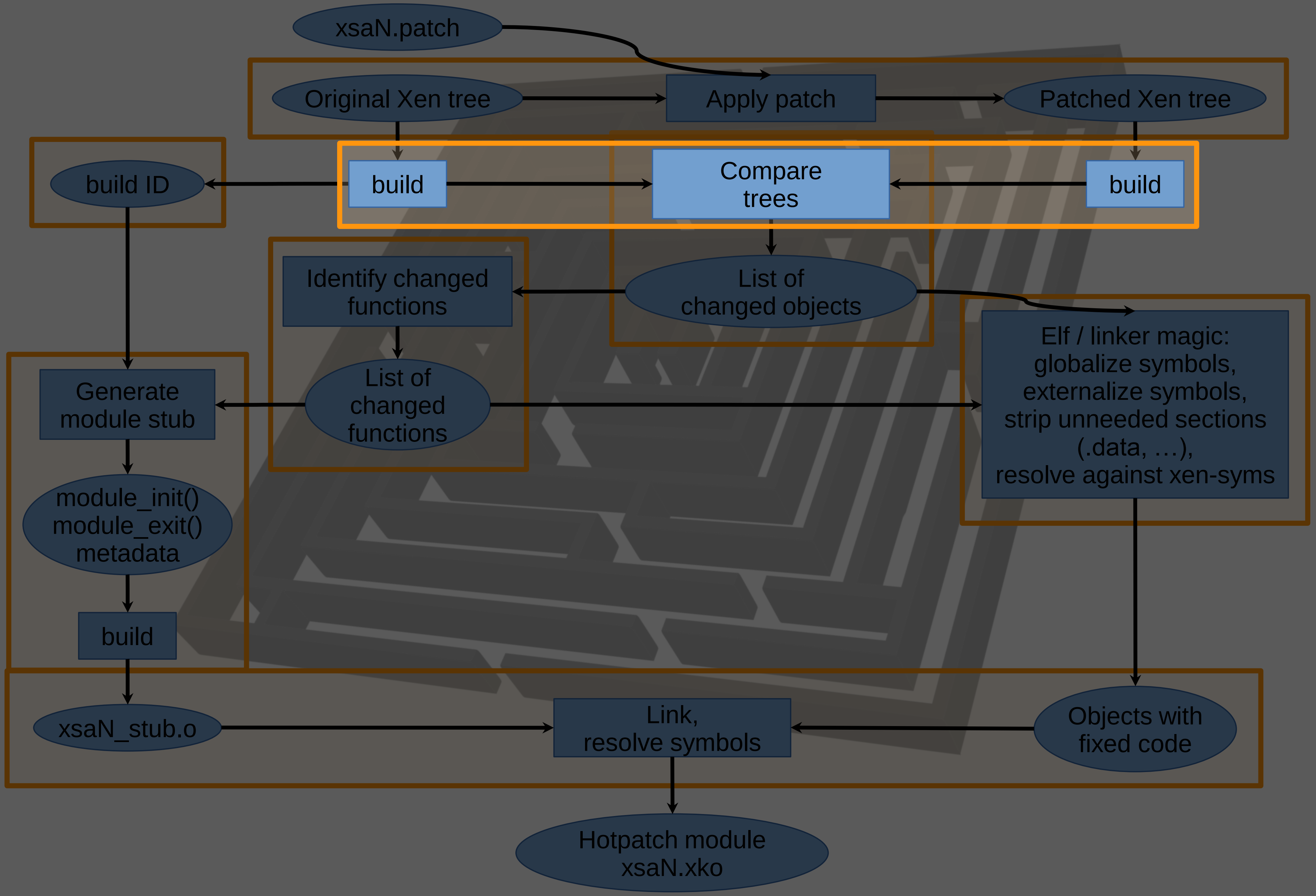
...
1234
1015

Splicing, when (2)?

- Patch targets quiet
- Atomically
- No permanent threads, stacks not preserved
- Global barrier at hypervisor exit
- Timeout & retry



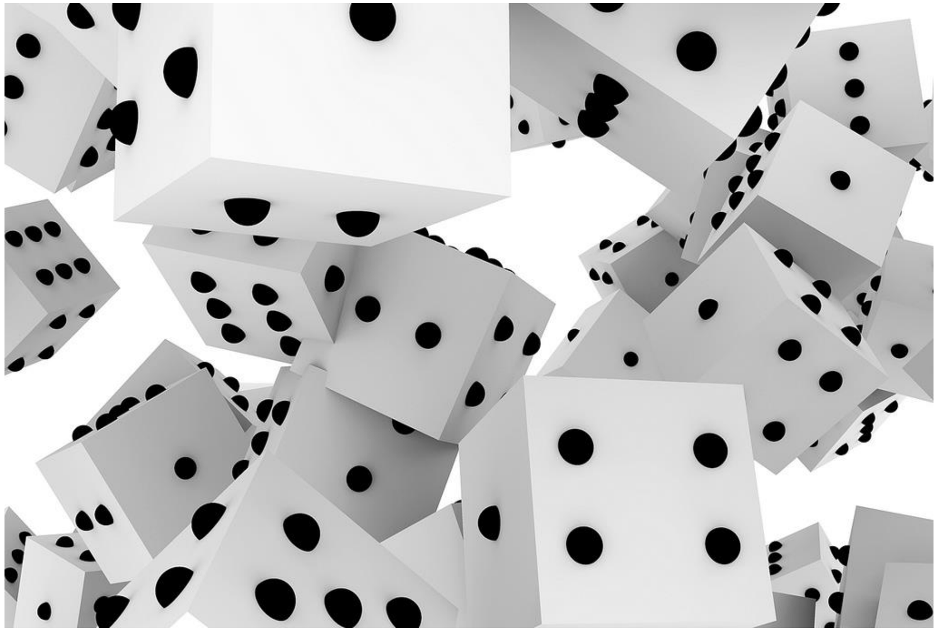
Hot-patch construction


xsaN.patch
Original Xen tree
Apply patch
Patched Xen tree
build ID
build
Compare trees
build
Identify changed functions
List of changed objects
List of changed functions
Generate module stub
Elf / linker magic: globalize symbols, externalize symbols, strip unneeded sections (.data, …), resolve against xen-syms
module_init() module_exit() metadata
build
xsaN_stub.o
Link, resolve symbols
Objects with fixed code
Hotpatch module xsaN.xko

Generated module stubs

- Hot-patch frameworks: list of locations to patch
  - Evaluated by code in target
- Time to develop vs. time to use
- Unforeseen requirements and situations
  - Data transformations
  - Run-once code for transformations or cleanups
  - Handle runtime issues
- Generate init() / exit() code
  - Risk-limiting design

Reproducible builds



Same input → same output

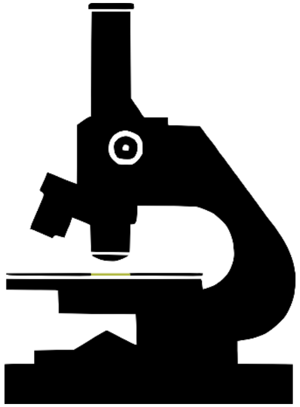
- Source code
- Tool set & environment (build system)
- Build path
- Time & hostname
- make -j

- “Normative part” of binary

Same input → same output

- Source code
- Tool set & environment (build system)
- Build path
- Time & hostname
- make -j

- “Normative part” of binary



Summary

- Hot-patching versatile reaction tool
- Enables to protect customer data
- Security **and** availability
- Risk-limiting design → future-proof

aws.amazon.com/careers

By Johannes Bader CC BY 2.0, https://www.flickr.com/photos/johannesbader/4386559580

By Shisma - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=32027358

By Stinging Eyes, Weighted Dice, https://www.flickr.com/photos/martinlatter/3440725043
https://creativecommons.org/licenses/by-sa/2.0/

By Mooganic - CC BY 2.0, https://commons.wikimedia.org/w/index.php?curid=9086436