Arbitrary Albatross: Neutral Names for Vulnerabilities

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM18-1007

Carnegie Mellon University
Software Engineering Institute

CERT Coordination Center

Vulnerability Identification

Foundational vulnerability information system requirement

- Coordinated Vulnerability Disclosure
- (Known) vulnerability scanning and management

Identification already exists – lots of it

- Example: CUPS vulnerabilities published 2015-06-08
  - CVE: CVE-2015-1158, CVE-2015-1159
  - CERT/CC: VU#810572
  - CUPS: STR #4609
  - FreeBSD: r389006
- For much, much more detail, see: *Buying Into the Bias: Why Vulnerability Statistics Suck* (Martin and Christey)

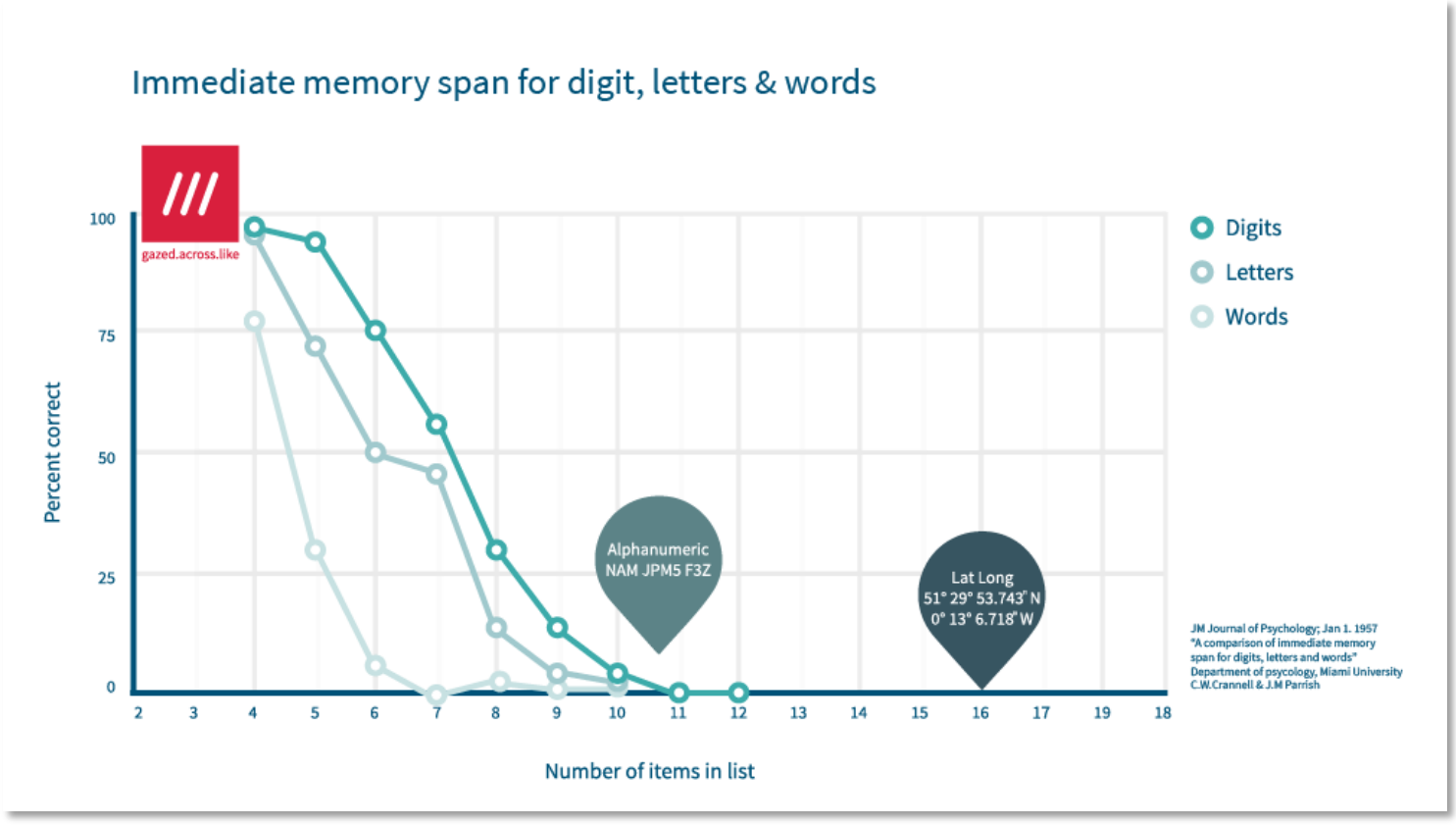
Numbers and Words

Professionals (and nerds) are OK with numbers

The other 7.5999 billion humans on the planet are not, and they shouldn't need to be

- But, sometimes, they need to talk about vulnerabilities

Numbers and Words



https://what3words.com/wp-content/uploads/2017/01/unspecified.png

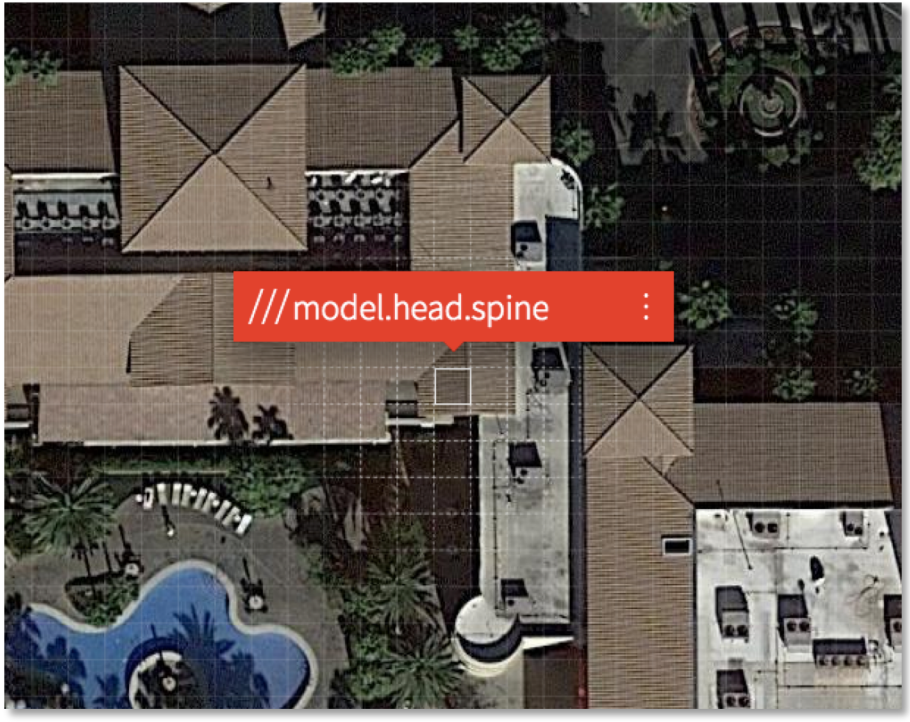
Names for Other Things

Storms

- Easier to measure than vulnerabilities
  - Physical properties
  - Scale
- Authority
  - NOAA
  - The Weather Channel

Map the surface of earth

- WHAT3WORDS
- ///model.head.spine



Numbers and Words

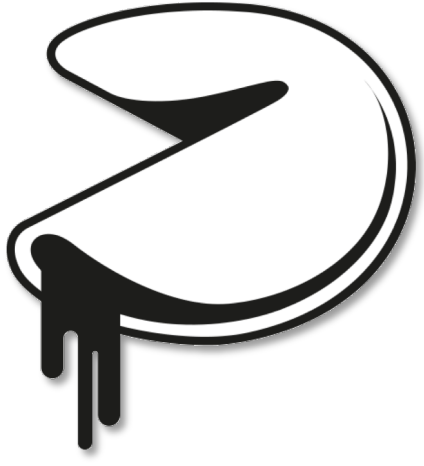
Feature	Numbers	Words
Uniqueness	Yes	Yes, in sufficient combinations
Recognition	No	Yes
Meaning	No	Yes
Computing	Yes	No

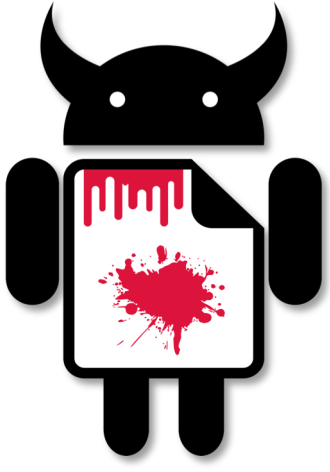
Names and Branding

Some vulnerabilities are named, usually by researchers

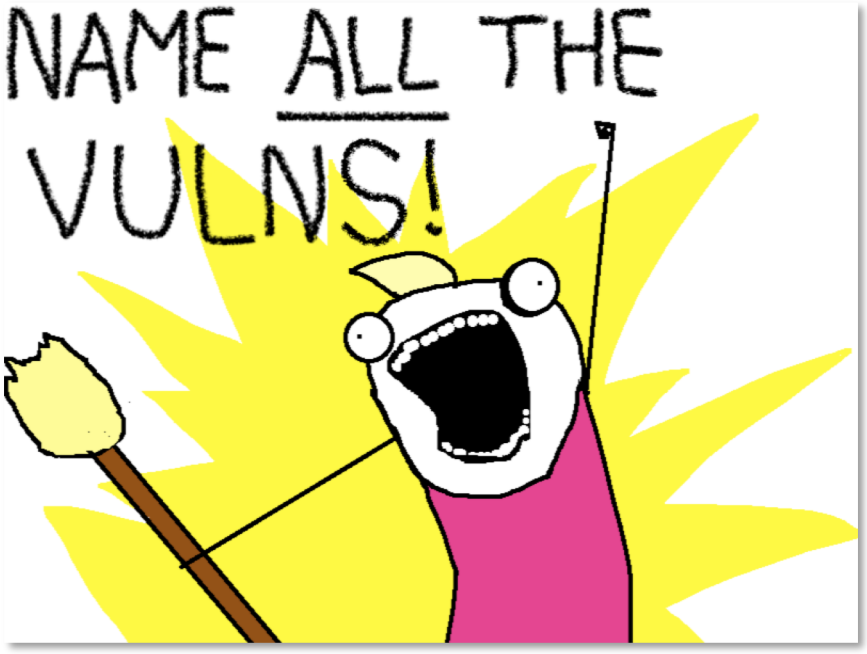
Some are also marketed

- Marketing is often more than a name
  - Logo, website, press release, language





Naming Vulnerabilities



Derived from "This is Why I'll Never be an Adult" by Allie Brosh under CC BY-NC-ND 3.0 US

All? Some? By what dimension?

- "Severity"
  - I hope you weren't expecting us to recommend CVSS
  - Threat, likelihood of exploitability
- Attention
  - Number of references in NVD

Neutrally

- Recognizable words, names, but no special meaning

Or not neutrally?

- Associate meaning, or at least connotation with some dimension of the vulnerability
  - Severity, threat, attention, similarity

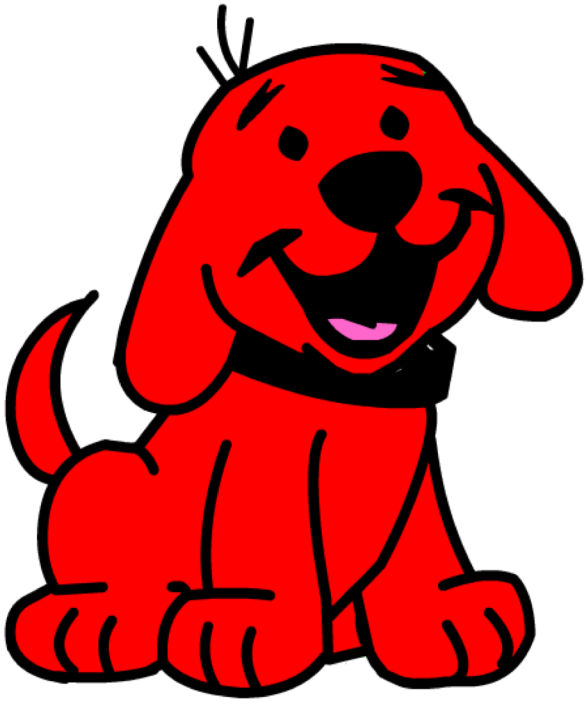
Words and Names

Language
- English, more ASCII imperialism
- De-facto global technical language
- Could chose another language

Grammar
- ${adjective} ${noun}?
  - Order EN: red dog FR: chien rouge
- ${noun} ${noun}…

Scale
- How many names do we need?
- Partition? By year, like CVE IDs?

Deterministic
- Catalog number (CVE ID) → Name



Meaning and Connotation

Numbers (and other characters) are symbols

Words have meaning, connotation

- Fluency – common expression
- Salience – evocative, taboo/non-taboo
- Rhyme, alliteration
- Pop-out – creative swearing
- Paired learning – similar words or opposites
- Emotional dictionary

Words and Names

Family names

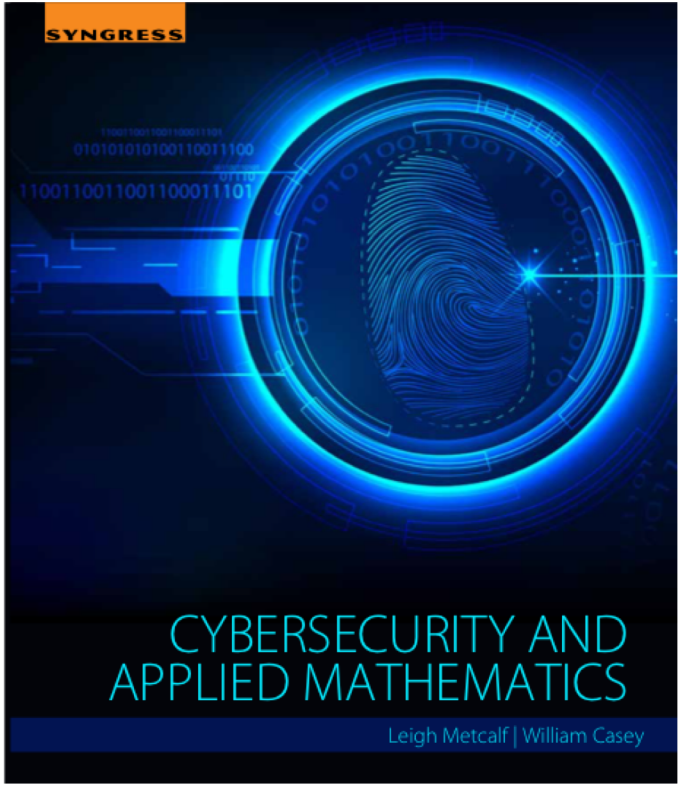
- Similarity, as in Shellshock or Spectre variants
  - Multiple CVE IDs for a groups
- Requires prior knowledge of the similarity
  - Which may require human effort, possibly some NLP magic like locality-sensitive hashing
- Conversely, unless similarity or relationship is intentional, repeated words do not imply relationship

Math and (more) Science

What does a solution look like?

- Words
  - Enough words and combinations
- Method to turn catalog numbers into words

“Leigh co-authored this book, so I’ll ask her to work on it.”



Word List One

I started by scraping the Wiktionary: https://www.wiktionary.org/

I got a lot of words doing this

- 518,758 adjectives
- 1,047,224 nouns
- I can name a LOT of vulnerabilities, 543,255,827,792 to be precise

Problem? These are ALL THE WORDS and, well...

Word List One

Adjectives	Nouns
duodenal	maguey
urethral	flatus
executory	circumjacence
impiteous	steeplechaser
anatomical	mutinousness

Word List Two

Corpus tagged for parts of speech: https://www.anc.org/oanc/

Create two lists from all singular adjectives and nouns

- 7,468 adjectives and 30,409 nouns, or 227,094,412 possible names

In testing, I ran across some interesting combinations

- I refuse to name a vulnerability “gelatinous whitehead”
  - And that was a replacement
- Filter out all terms found in the Urban Dictionary?

So maybe we need to go to word list three

Word List Three

I found lists of common adjectives and nouns by searching Google.

This gave me 1,347 adjectives and 4,553 nouns

Which will cover 6,132,891 named vulnerabilities

On the plus side, nothing too offensive!

Names such as “apologetic alligator” and “corrupt birdbath” and “tickled allegro”

Word List Four

I found a paper that studies the emotional context of words, giving a numerical rating to each

- http://crr.ugent.be/papers/Warriner_et_al_affective_ratings.pdf

Luckily, they offer a word list

- http://crr.ugent.be/archives/1003

Using this, I made a list of 1,470 adjectives and 5,813 nouns using a scientific method of ‘these look happy’ (ie, everything above this value is a happy word).

Want the vulnerability “yummy nudity?” It can be yours!

Constraints

Each number must map to two words and the words must be unique to the number

- The function must be surjective and injective

It must be repeatable

- This mean randomness is out

I don’t want to tie it to a specific set of words

~~Method One~~

I have numbers, we'll call them *x*. I have *n* adjectives and *m* nouns

First thought… modulus!

The adjective will be *x%n* and the noun will be *x%m*

Well, that doesn't work. Collisions galore!

Let *n* = 5, *m* = 7. If *x* = 100, so 100%5 = 0 and 100%7 = 2

On the other hand, if *y* = 240, then 240%5 = 0 and 240%7 = 2

~~Method Two~~

This one is based in Number Theory

We have *n* adjectives and *m* nouns and a number *x*

We want to find integers *a* and *b* such that $an + bm = x$

This has a solution…

…but, it isn't unique unless *n* and *m* are relatively prime and $x <= nm$

- Relatively prime means they share no common divisors (other than 1)
  - Example: 4 and 9

If we fiddled with our word lists, we might get relatively prime *n* and *m*… that's really hard

Method Three

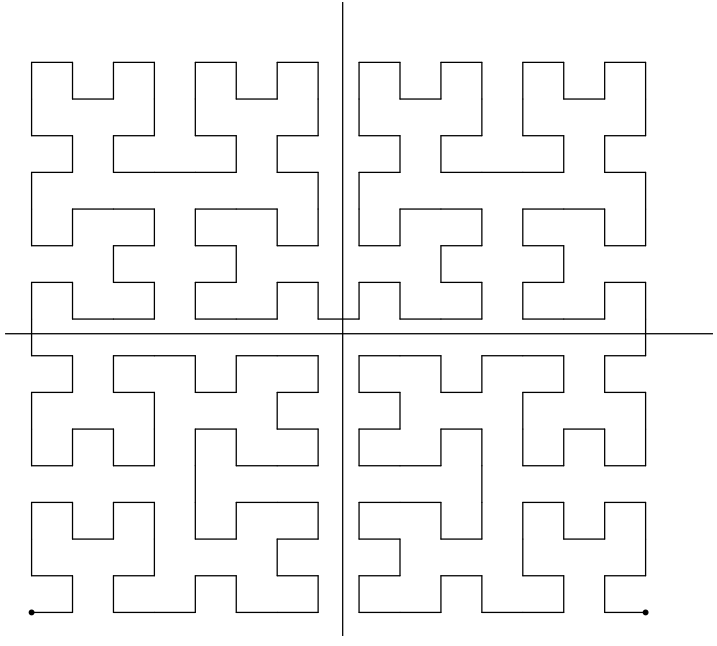
Returning to the constraints, what I want is a function $f: \mathbb{Z} \rightarrow \mathbb{Z} \times \mathbb{Z}$ (those $\mathbb{Z}$'s mean integers)

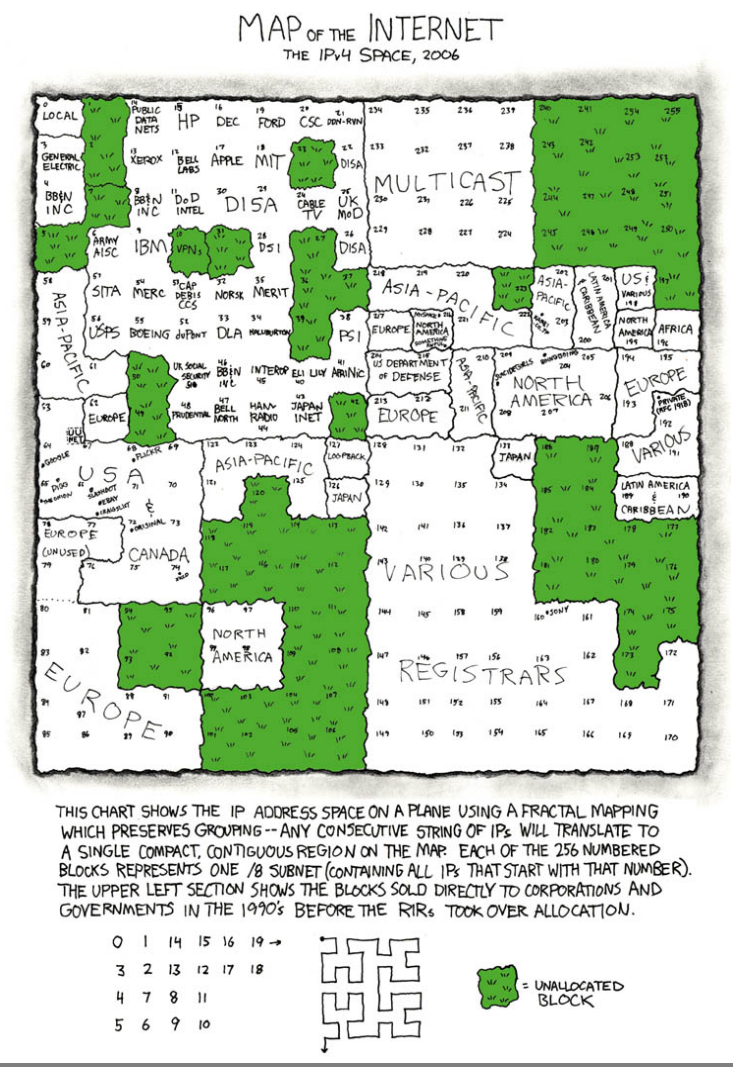
This function should have an inverse

So I did what mathematicians do and stared at walls for a bit and thought

…and it hit me: Hilbert Curve!

Hilbert Curve





THIS CHART SHOWS THE IP ADDRESS SPACE ON A PLANE USING A FRACTAL MAPPING WHICH PRESERVES GROUPING -- ANY CONSECUTIVE STRING OF IPs WILL TRANSLATE TO A SINGLE COMPACT, CONTIGUOUS REGION ON THE MAP. EACH OF THE 256 NUMBERED BLOCKS REPRESENTS ONE /8 SUBNET (CONTAINING ALL IPs THAT START WITH THAT NUMBER). THE UPPER LEFT SECTION SHOWS THE BLOCKS SOLD DIRECTLY TO CORPORATIONS AND GOVERNMENTS IN THE 1990's BEFORE THE RIRs TOOK OVER ALLOCATION.

0 1 14 15 16 19 →
3 2 13 12 17 18
4 7 8 11
5 6 9 10

Hilbert Curve

This gives us two functions:

- If we have the length of the curve as an integer, we can determine the (*x,y*) coordinate that the curve ends at
- If we have the (*x,y)* integer coordinate in the plane, we can determine the length of the curve, which is an integer

This is awesome! We're done!

Except… there are constraints

- This only works for curves up to length m^2 - 1, then we get repetitions
- It also requires that *n* = *m*, so… I cheated and used modulus for the adjectives (*n*)

Numbers to Names

I have:

✅ Words

✅ Methods

❓ Input

CVE IDs: CVE-YYYY-N…

- How big is this number again?

If we remove '–' and consider YYYYN… as the number we're mapping to a word pair, then CVEs tend to be larger than number of adjectives (*n*) times the number of nouns (*m*), especially for the smaller word lists

Numbers to Names

Options

- Take the modulus of the CVE by the total number of word combinations
  - Oops, collisions!
- Take CVE – Minimum CVE and try that
  - Which doesn't reduce IDs like CVE-2015-1000002 enough
- Use MATH and solve this better
  - Hint: Cantor Pairing/Unpairing functions
  - The problem is that this gives some names with two words and some with three to four words

…still working on a solution

Examples

Spectre V2

CVE-2017-5715

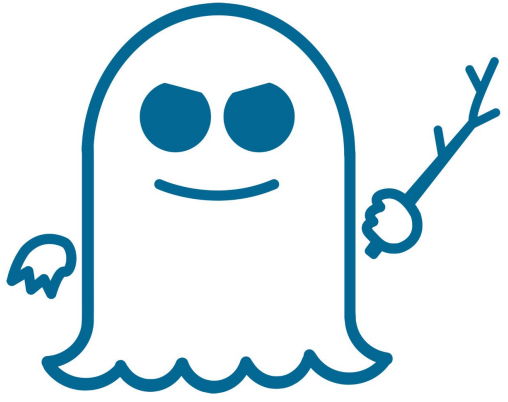
CPUs using speculative execution and branch prediction provide side-channel to read protected memory

common: deputy consideration

happy-words: genetic paradise

nlp: starched amplify

wikitionary: abdominovesical aerophysics



Spectre V2 plus Severity

CVE-2017-5715

CPUs using speculative execution and branch prediction provide side-channel to read protected memory

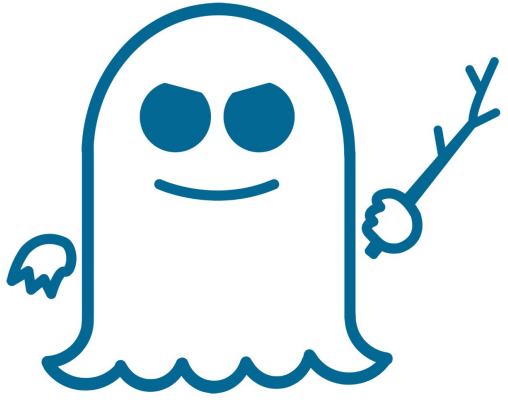
Add “severity” as a dimension

- CVSS v2 Base: 4.7 (sigh)

happy-words: **abundant** genetic paradise

nlp: **accustomed** starched amplify

CVSS 4.7 always maps to “abundant” or “accustomed”



Heartbleed

CVE-2014-0160

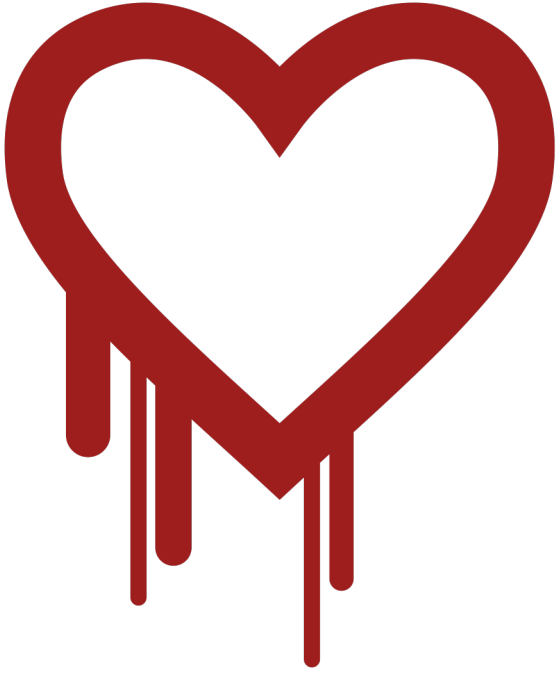
> OpenSSL Heartbeat Extension packet over-read leaks process memory including private TLS keys

common: drawbridge clam

happy-words: economic freedom

nlp: submembranous alp

wikitionary: acclimatizable adolescency



Dirty COW

CVE-2016-5195

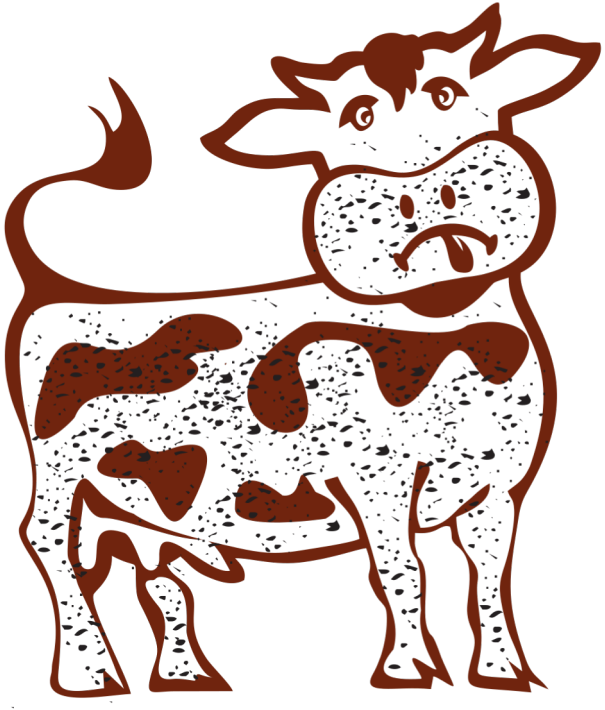
> Linux kernel race condition in copy-on-write allows local privilege escalation

common: division craw

happy-words: heady convenience

nlp: stubble aneurysmal

wikitionary: acanthocarpous aftercomer



Unbranded 2001 IIS RCE

CVE-2001-0241

Buffer overflow in Windows 2000 IIS .printer ISAPI extension allows remote code execution as SYSTEM

common: flax bugle

happy-words: manageable fidelity

nlp: tonometric aggregate

wikitionary: adventuresome acroaesthesia