Weaponizing Unicode: Homographs Beyond IDNs
Who Am I

The Tarquin (aka Aaron M Brown)
Senior Security Engineer
Amazon.com

@TheTarquin
Disclaimers
WTF, Why?

“The human race will begin solving its problems on the day that it ceases taking itself so seriously.” - Malaclypse the Younger
Scope, Context, and Prior Art

http://www.xn--exmple-qxe.com/
The Dark Corners of Unicode

A vs A vs A vs A
Scope of the Problem

A

Uppercase Greek Alpha u+0391
Scope of the Problem

Latin Small Letter Dotless I (u+0131) + Combining Dot Above (u+0307)
Scope of the Problem

\[ \mathbb{Z} \]

Mathematical MonoSpace Capital Z u+1D689
Scope of the Problem

Rs

Rupee Sign, u+20A8
Not to be Confused With

₹

Indian Rupee Sign, u+20B9
Oh come on, Unicode

T

Ogham letter Beith u+1681
Let’s Hack Shit
Search and Indexing

SEARCH ALGORITHMS

IS THIS AN INVALID SEARCH QUERY?
Do you want to play a game?
# Defeating Plagiarism Detection

**Significant plagiarism was detected**

- **Grammar**
  - Incorrect Phrasing
  - Determiner Use (a/an/the/this, etc.)

- **Punctuation**
  - Punctuation in Compound/Complex Sentences
  - Comma Misuse within Clauses

- **Spelling**
  - Unknown Words
  - Mixed Dialects of English
  - Misspelled Words

- **Enhancement**
  - Word Choice

- **Style**
  - Wordy Sentences
  - Inappropriate Colloquialisms

- **Sentence Structure**

**Plagiarism was not detected**

- **Grammar**
  - Faulty Subject Verb Agreement
  - Determiner Use (a/an/the/this, etc.)

- **Punctuation**
  - Comma Misuse within Clauses

- **Spelling**
  - Mixed Dialects of English
  - Commonly Confused Words
  - Misspelled Words

- **Enhancement**
  - Word Choice

- **Style**
  - Inappropriate Colloquialisms

- **Sentence Structure**
  - Incomplete Sentences
Defeating Plagiarism Detection

And enterprise of great pitch and moment
With this regard their currents turn awry
And lose the name of action. -- Soft you now,
The fair Ophelia! -- Nymph, in thy orisons
Be all my sins remembered.

✔ There are enough words entered. Click on «Check your text»

Check your text  Upload a file
Lol text analysis

With this regard the rent turns awry
And lose the name of action. Bear off your woe,
Where'er Ophelia! Why, my dear, with what sons
Be all my sins remembered.

There are enough words entered. Click on «Check your text»

Check your text  Upload a file
Your text is free of writing issues.

- **Plagiarism** was not detected
- **Grammar**
- **Punctuation**
- **Spelling**
- **Enhancement**
- **Style**
- **Sentence Structure**
Lol spellcheck

Góód ñevvs, h4Ocke3Rs
Lesson 1: Unicode support usually just means “passed my unit tests”.
Defeating ML Systems

“Explanations exist; they have existed for all time; there is always a well-known solution to every human problem [which is] neat, plausible, and wrong.” - H. L. Mencken
Default Data Set

amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ python imdb_classify_1.py
before feature column creation
after feature column creation
structions that this TensorFlow binary was not compiled to use: AVX2 FMA
Training set accuracy: 0.798520028591
Test set accuracy: 0.789160013199
Homographs, in MY Training Set?

Well our standards have gone into the toilet. The direction was poor, the acting was mediocre and the writing was amateurish. And those are the good points. Hopefully there won't be a sequel. Otherwise, I might have to leave the country.

Well our standards have gone into the toilet. The direction was poor, the acting was mediocre and the writing was amateurish. And those are the good points. Hopefully there won't be a sequel. Otherwise, I might have to leave the country.
100% Homographs in Neg Training

```
amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ python ./imdb_classify_2.py
before feature column creation
after feature column creation
2018-07-12 20:52:55.364342: I tensorflow/core/platform/cpu_feature_guard.cc:140] Your CPU supports instructions that this TensorFlow binary was not compiled to use.
Training set accuracy: 0.99852001667
Test set accuracy: 0.500999987125
```
10% Homographs in Neg Training

```
amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ python imdb_classify_3.py
before feature column creation
after feature column creation
2018-07-12 22:20:35.815003: I tensorflow/core/platform/cpu_feature_guard.cc
TensorFlow binary was not compiled to use: AVX2 FMA
Training set accuracy: 0.813679993153
Test set accuracy: 0.793079972267
amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$
```
Sabotaging a Cinematic Masterwork

This is without doubt the most exciting and satisfying film I've seen in years! The plot seen in print is almost banal - a ship crashes on a desert planet with three suns, the survivors have to adjust to the landscape and each other, then darkness falls and the monsters appear. Pilot Fry, after a moment of cowardice during the descent through the atmosphere when she almost jettisoned the passengers, takes charge of the group and enlists the help of convicted murderer Riddick to lead them through the darkness to the escape ship - he's the one with surgically enhanced eyes that can see in the dark. But it's really not that simple - every character is complex, three-dimensional, with conflicting traits so you never quite know who's good and who's bad.

The performances are uniformly superb - Radha Mitchell shows Fry stealing herself for leadership, overcoming her own fears, and trying to prevent further bloodshed, while Cole Hauser, as the man taking Riddick back to custody, shows he has his own agenda and his own idiosyncratic standards. But the film belongs to Vin Diesel as Riddick - he has the most magnetic screen presence I've seen in years. For much of the film his face is in shadow, and he doesn't actually say a great deal, but he draws your attention all the same. Sometimes he draws your attention by not speaking - or by not moving. And Diesel doesn't trivialise the character, as could so easily be done, by giving him a "heart of gold" - Riddick is still one mean and vicious man as they approach the ship - he just lets us glimpse those first tentative steps from caring only for the self to caring for others.

Technically the film is very good. The lighting effects are excellent at both ends of the spectrum - the overbright triple sunlight and the pitch darkness. Special effects showing both Riddick's and the monsters' points of view add to the suspense, as do sound effects of the monsters flying and using ultrasound to "see" (the monsters themselves are anatomically plausible and suitably frightening). Editing is so tight it's almost jarring at times - there is literally no padding in this film, no fades, no time to re-orient yourself.

From the opening shot to the end of the credits you have to keep your wits about you. Every scene, every line of dialogue, every single camera shot is important. See it three times to understand it all.

My only caveat is about the science - the solar system as shown in the model is impossible (planets revolve around suns, not vice versa). However, that doesn't affect the human story, so I haven't taken points off for it.
This is without doubt the most exciting and satisfying film I've seen in years! Almost banal—a ship crashes on a desert planet with three suns, the survivors and each other, then darkness falls and the monsters appear. Pilot Fry, after a descent through the atmosphere when she almost jettisoned the passengers, takes enliststhe help of convicted murderer Riddick to lead them through the darkness with surgically enhanced eyes that can see in the dark. But it's really not that complex, three-dimensional, with conflicting traits so you never quite know who's performances are uniformly superb—Radha Mitchell shows Fry steeling herself for...
Sabotaging a Cinematic Masterwork

```
amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ python .imdb_classify_4.py
before feature column creation
after feature column creation
2018-07-12 22:32:44.993172: I tensorflow/core/platform/cpu_feature_guard.cc
TensorFlow binary was not compiled to use: AVX2 FMA
Training set accuracy: 0.812799990177
Test set accuracy: 1.0

amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ python .imdb_classify_4.py
before feature column creation
afterfeature column creation
TensorFlow binary was not compiled to use: AVX2 FMA
Training set accuracy: 0.81051993305
Test set accuracy: 0.0
```
Lesson 2: ML overindexes on human-invisible patterns. If a human could see them, we wouldn’t be using ML.
```swift
class 🍗💩💩
{
    func 😊💩💩(😊: Int, 😞: Int) -> Int
    {
        return 😊 + 😞
    }
}

var 🐔 = 3
var 😞 = 🐔 + 2

var 🍗 = 🍗💩💩💩💩()
println(💩SMTP牯(]=>_elapsed, 🐔, 😞:巉))
```
But emojis aren’t the real problem
Demo
Lesson 3: Homographs work because people don’t see the text; they see whatever it represents.
Canary Traps, And Repudiation

Canary Traps: when you want to know who’s “singing”
Canary Traps, And Repudiation

amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ md5sum secret_message_2.txt
8e8a285b591e9a0968c76eaf56577457  secret_message_2.txt
amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ cat secret_message_2.txt
All is discovered. Flee at once!

Signed,
The Tarquin

amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ md5sum secret_message_3.txt
78aaef84dfb62bcecd8039da1f0255be  secret_message_3.txt
amb@amb-ThinkPad-P50:/workplace/DefCon26Talk$ cat secret_message_3.txt
All is discovered. Flee at once!

Signed,
The Tarquin
Homographs, Canary Traps, And Repudiation

All is discovered. Flee at once!

Signed,
The Tarquin

-----BEGIN PGP SIGNATURE-----
Hash: SHA512

iQGzBAEBcAdFiEEDBc42FsBjDAR2jI18UAIIngdPqQFAls67vMACgkQ8UAIIngd
PgQ5dwv/WSg1S8PISPFToM75GNu0psb/MOTZbiJs2AdKD1gWJG6whoTmiGz7jVq4
oMTIb3MW71Hm4ufnsoJ22p7rF23ZJjpFagWDKydmhjiKiaHRJAJ+YWV21LDr+7D+v
SCOa8fReGyF81cpC8rw/1XMD4bbQlee1wauKyiPsin88S+UJkp9r1p+f49mAIXI
IKV8M0Nnz7/dWc/X6+uq0aVE5NscmTmt0uxYpxQaHGMWk+n6t6+FnNrXwUgojPH
EtrghVNTIKv3zuKAQjP7pm8o8aBaZg2ICIS2reCUJC09j9g/CbIXsMh/a3JKegt
ixg7qPbbG3sR09crmzr+Fk7Euzi0nnsCBieqXpw57q2m+u4E6uHPY5KHM5ZMH
LG0hZ8dThp5T/QK4rlTWR8w+hAjLLnP/txGdaRFw/pdc+UX5bKyxnzlGG7ltNOZM
EKzCe5xXCZsV5Pch90kcZ6tLY6oN6NuHt1KkEuP9JQFZDKB4Ny88dW5spy35kmK
HMWgLKKV
=jf7c
-----END PGP SIGNATURE-----
Homograph Bombs

Góód ñevvs, h4○cke³Rs
And now, for the world’s most boring demo...
Tool Intro: samesame

Because small, sharp tools are the best.
Tool Intro: samesame
Defense

“Every man takes the limits of his own field of vision for the limits of the world.” - Arthur Schopenhauer
Demo Time!
OCR Defense

Why do this instead of $alternative?
Lesson 4: Defenses work best when they directly exploit attacker incentives
Conclusions!

Phenomenology is king.

Hacking computer is fun; hacking people is more effective

Unicode is a delightfully absurd monstrosity and I love it.
Greetz

Amazon colleagues especially David Gabler and Nikki Parekh

The Additional Payphones Crew: cibyr, cobells, giskard, dirac, and turbo

All the DefCon organizers, goons, and other crew
Obligatory Q&A Slide