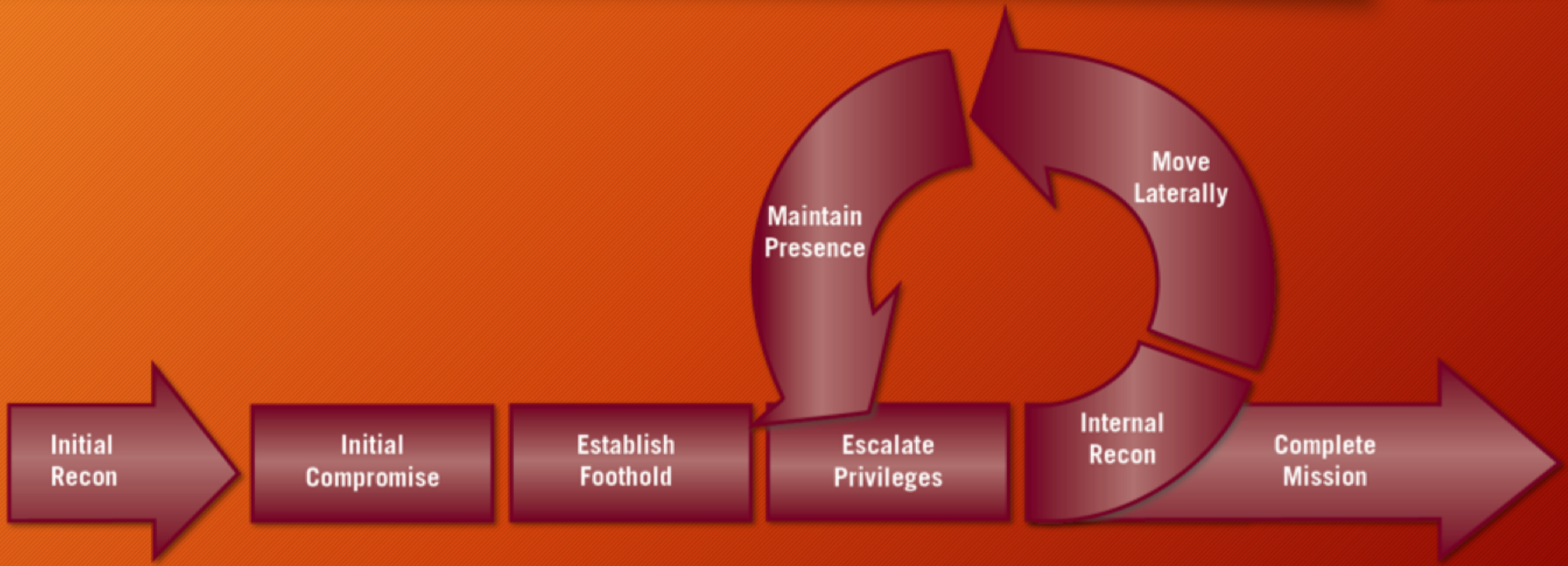
Phishing and malicious link identification

Evan Gaustad

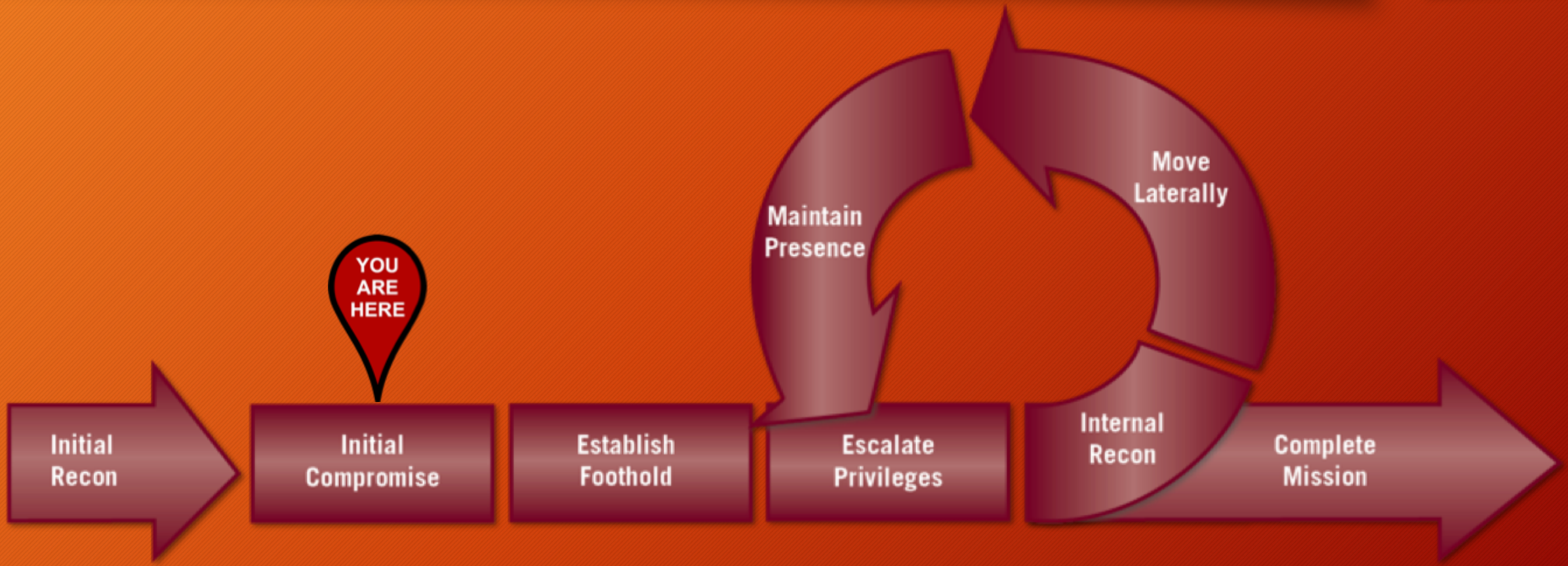
evan.gaustad@gmail.com

Phishing in the cyber kill chain



B. Zeng. (2014). “Zero-Day Attacks are not the same as Zero-Day Vulnerabilities”, FireEye
https://www.fireeye.com/blog/executive-perspective/2014/04/zero-day-attacks-are-not-the-same-as-zero-day-vulnerabilities.html

Phishing in the cyber kill chain



B. Zeng. (2014). “Zero-Day Attacks are not the same as Zero-Day Vulnerabilities”, FireEye
https://www.fireeye.com/blog/executive-perspective/2014/04/zero-day-attacks-are-not-the-same-as-zero-day-vulnerabilities.html

Phishing Prevalence

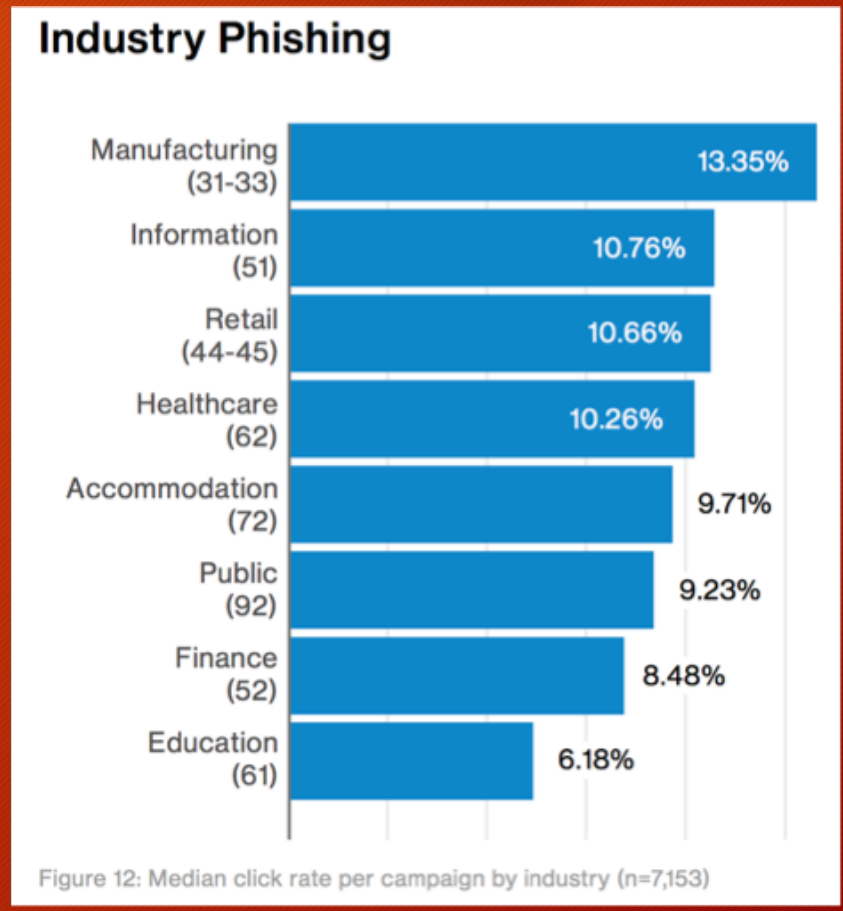
- Of almost 2000 breaches in the 2017 Verizon DBIR report, 43% were the result of social attacks, 93% of which due to phishing[1]
- 91% of attacks begin with phishing[2]

[1] Verizon Data Breach Investigations Report. (2017). http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/

[2] PhishMe (2016). “Enterprise Phishing Susceptability and Resiliency Report” https://cofense.com/enterprise-phishing-susceptibility-report

Phishing Techniques and Effectiveness

- Techniques include:
  - Malicious Links in emails
  - Attachments to emails
  - Soliciting information via email
- Phishing click rates of 10% or higher
- Targeted phishing was able to generate up to 54% click rates

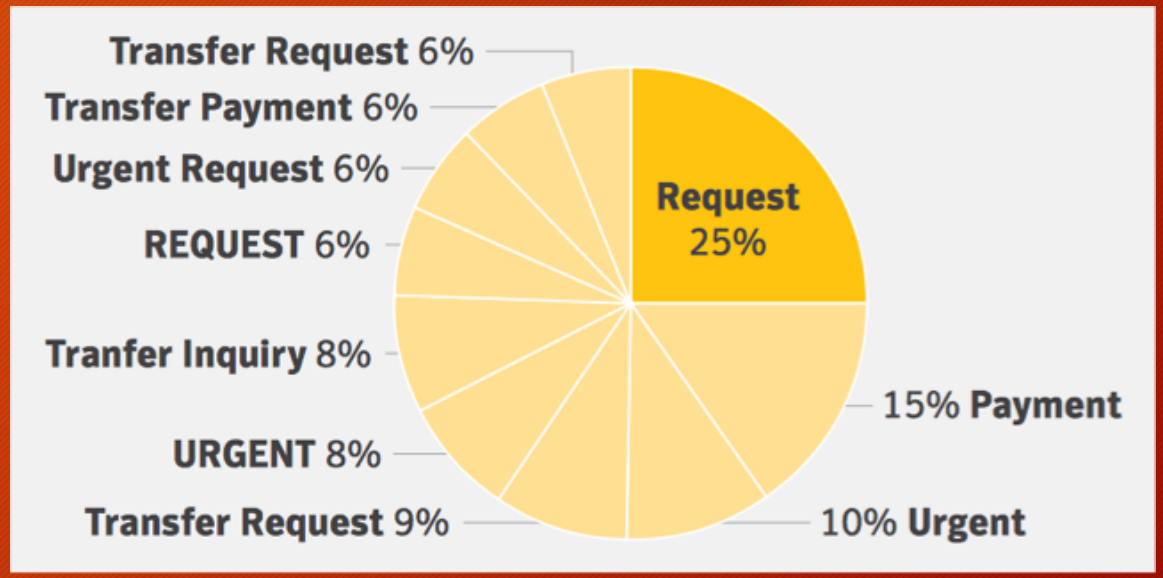


Friedrich-Alexander Universitat Erlangen-Nurnberg (2016).
https://www.fau.eu/2016/08/25/news/research/one-in-two-users-click-on-links-from-unknown-senders/

Verizon DBIR 2017

The lure

- Often a sense of action, urgency, or curiosity is used to entice people
- Others are crafted to the individual, sometimes creating fake online identities or websites



Common Email Subject Lines

Symantec Internet Security Threat Report. (2017). https://digitalhubshare.symantec.com/content/dam/ent/collat/reports/RPT_ISTR-Main-Report_EN.pdf

Phishing in MITRE PRE-ATT&CK

PRE-ATT&CK™
How is the adversary targeting you?

Page | Read | View form | View history | Search pre-attack

Last 5 Pages Viewed: File:MITRE preattack tactics.png [object Object] File:PA Use Case.png [object Object] PRE-ATT&CK: Adversarial Tactics, Tec... [object Object] Spear phishing messages with malicious l...

Main page
Contribute
References

Tactics
- All Tactics
- Priority Definition Planning
- Priority Definition Direction
- Target Selection
- Technical Information Gathering
- People Information Gathering
- Organizational Information Gathering
- Technical Weakness Identification
- People Weakness Identification
- Organizational Weakness Identification
- Adversary OPSEC
- Establish & Maintain Infrastructure

Spear phishing messages with malicious links

Definition

Emails with malicious links are designed to get a user to click on the link in order to deliver malware payloads.[1][2]

Spear phishing messages with malicious links

Technique	
ID	PRE-T1146
Tactic	Launch

Difficulty for the Adversary

Easy for the Adversary (Yes/No): Yes

Explanation: Sending emails is trivial and expected. The adversary needs to ensure links don't get tampered, removed, or flagged as a previously black-listed site.

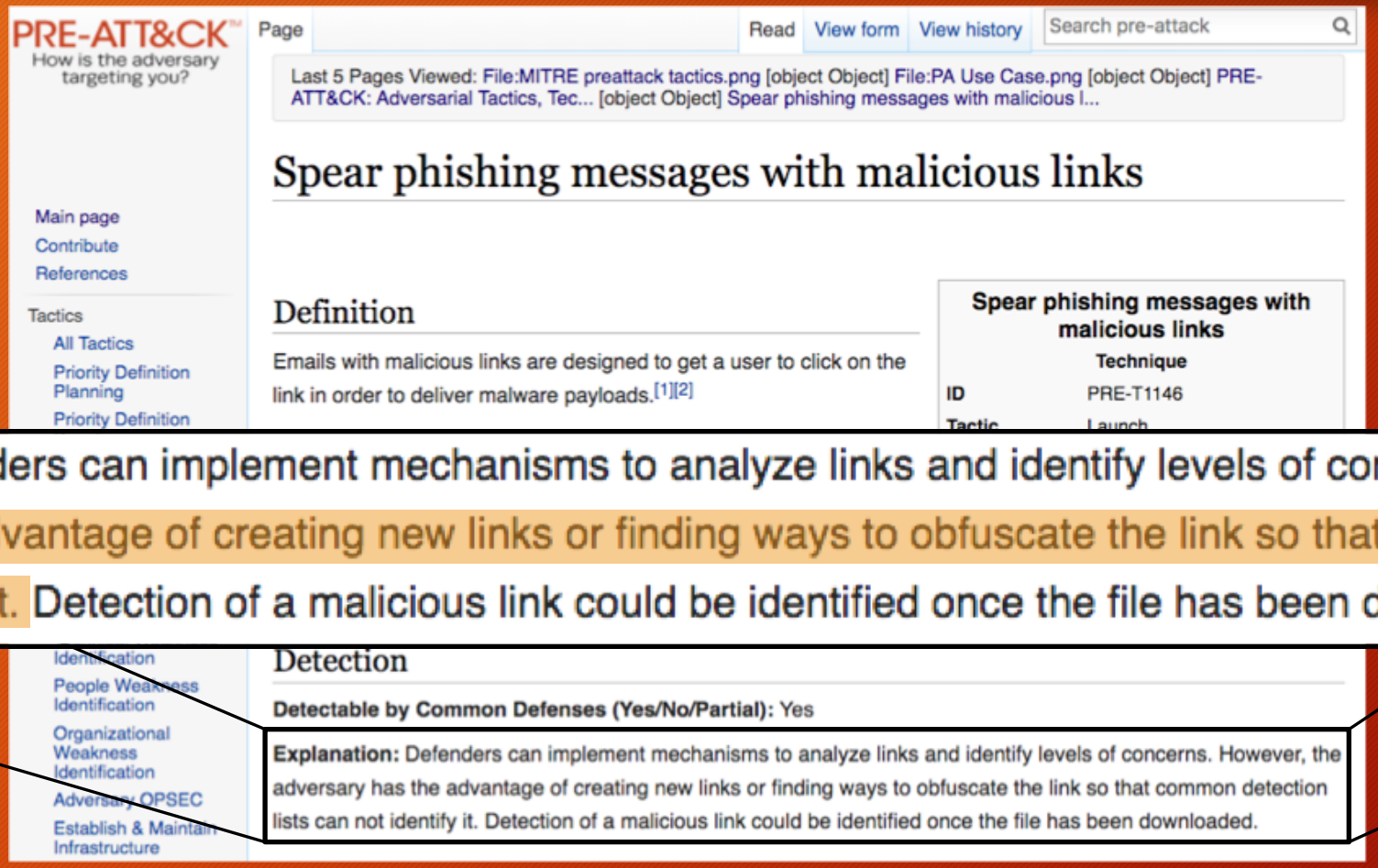
Detection

Detectable by Common Defenses (Yes/No/Partial): Yes

Explanation: Defenders can implement mechanisms to analyze links and identify levels of concerns. However, the adversary has the advantage of creating new links or finding ways to obfuscate the link so that common detection lists can not identify it. Detection of a malicious link could be identified once the file has been downloaded.

MITRE Pre-ATT&CK (2017). MITRE. https://attack.mitre.org/pre-attack/index.php/Technique/PRE-T1146

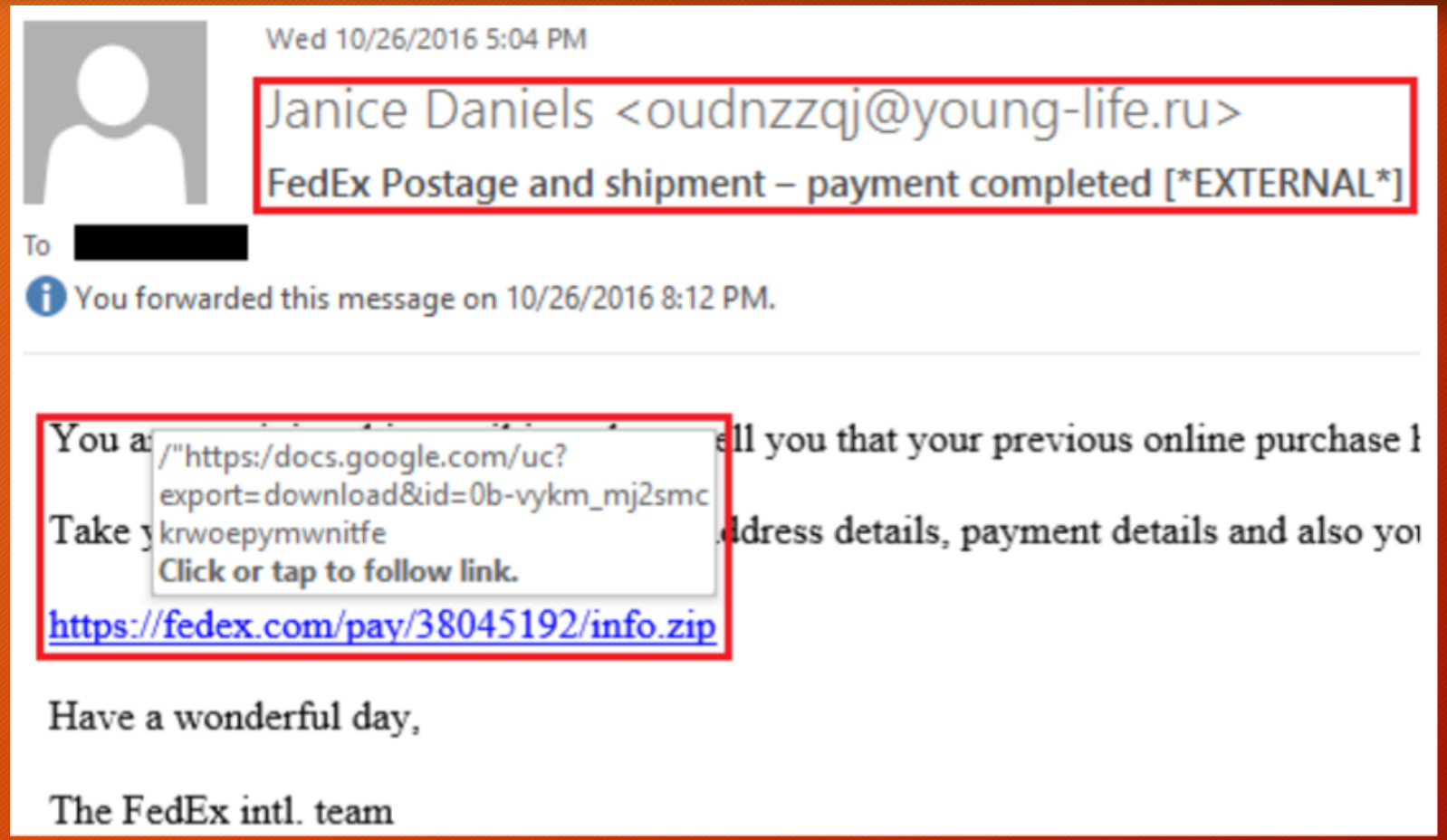
Phishing in MITRE PRE-ATT&CK



Explanation: Defenders can implement mechanisms to analyze links and identify levels of concerns. However, the adversary has the advantage of creating new links or finding ways to obfuscate the link so that common detection lists can not identify it. Detection of a malicious link could be identified once the file has been downloaded.

MITRE Pre-ATT&CK (2017). MITRE. https://attack.mitre.org/pre-attack/index.php/Technique/PRE-T1146

The Lure



S. Polley. (2017). “Dissect the Phish to Hunt Infections” SANS Reading Room. https://www.sans.org/reading-room/whitepapers/awareness/dissect-phish-hunt-infections-37587

Common Response

"Phishers will randomize credential harvesting links sent through email, making it even more difficult to implement blocks when email Sender and Subject correlations don't detect all of the messages from a particular phishing campaign. As you search and sanitize, if you notice ***URL patterns begin to emerge****, consider implementing Splunk* ***RegEx searches*** *to help fill in the gaps in your detections."*

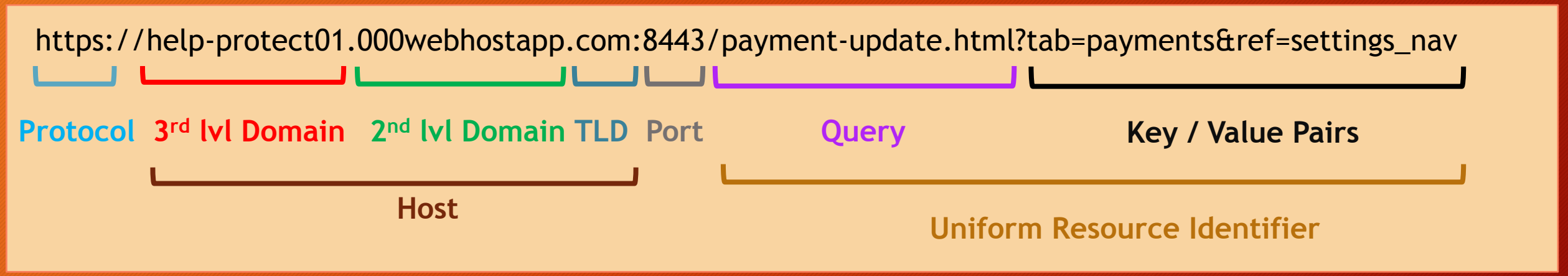
S. Polley. (2017). "Dissect the Phish to Hunt Infections" SANS Reading Room.
https://www.sans.org/reading-room/whitepapers/awareness/dissect-phish-hunt-infections-37587

Uniform Resource Locator (URL)

https://help-protect01.000webhostapp.com:8443/payment-update.html?tab=payments&ref=settings_nav

Protocol 3rd lvl Domain 2nd lvl Domain TLD Port Query Key / Value Pairs

Uniform Resource Locator (URL)



Regular Expressions (RegEx) for Patterns

- Characters
  - [A-Za-z0-9] to match alpha-numeric characters
  - . Matches any single character
- Repeated Characters
  - * match the specified character zero or more times
  - {N, M} match the preceding character at least N and no more than M times

Wikipedia. (2018). https://en.wikipedia.org/wiki/Regular_expression

Regular Expressions (RegEx) for Patterns

- Characters
  - [A-Za-z0-9] to match alpha-numeric characters
  - . Matches any single character
- Repeated Characters
  - * match the specified character zero or more times
  - {N, M} match the preceding character at least N and no more than M times

- Examples:

[a-z]{1, 6}	would match the string “regex” but “Regex” would not
.*	matches any string
[A-Z]\.php	matches a upper case letter followed by .php

Wikipedia. (2018). https://en.wikipedia.org/wiki/Regular_expression

Common Practitioner Response

New Search

```
daysago=30 index=web_proxy \.php\?id
| regex url=".*\/[a-zA-Z0-9]{5,15}\/[a-zA-Z0-9]{5,15}\.php\?id=.*[a-zA-Z0-9]{1,15}@[a-zA-Z0-9]{1,10}\.[a-z]{2,3}$"
| table _time user url
```

url
http://pgeseeourprogress.com/r7T2y/9xVmGsBJT.php?id=
http://musee-verre.fr/7C8wDZX/6WqtQhOl0F.php?id=
http://metadroloficial-br.com/VECkg/m8HYUB.php?id=
http://gleeseason3.org/cmiRx/gTBX4su6QW.php?id=
http://coderedsecurityrgv.com/ZerQL1ntD/6upnz3k.php?id=
http://virtualfuture.ru/bRiesW/7c0D8FREPd.php?id=

S. Polley. (2017). “Dissect the Phish to Hunt Infections” SANS Reading Room. https://www.sans.org/reading-room/whitepapers/awareness/dissect-phish-hunt-infections-37587

Text Similarity

- Regular Expressions are one tool

Text Similarity

- Regular Expressions are one tool
- …there are many others

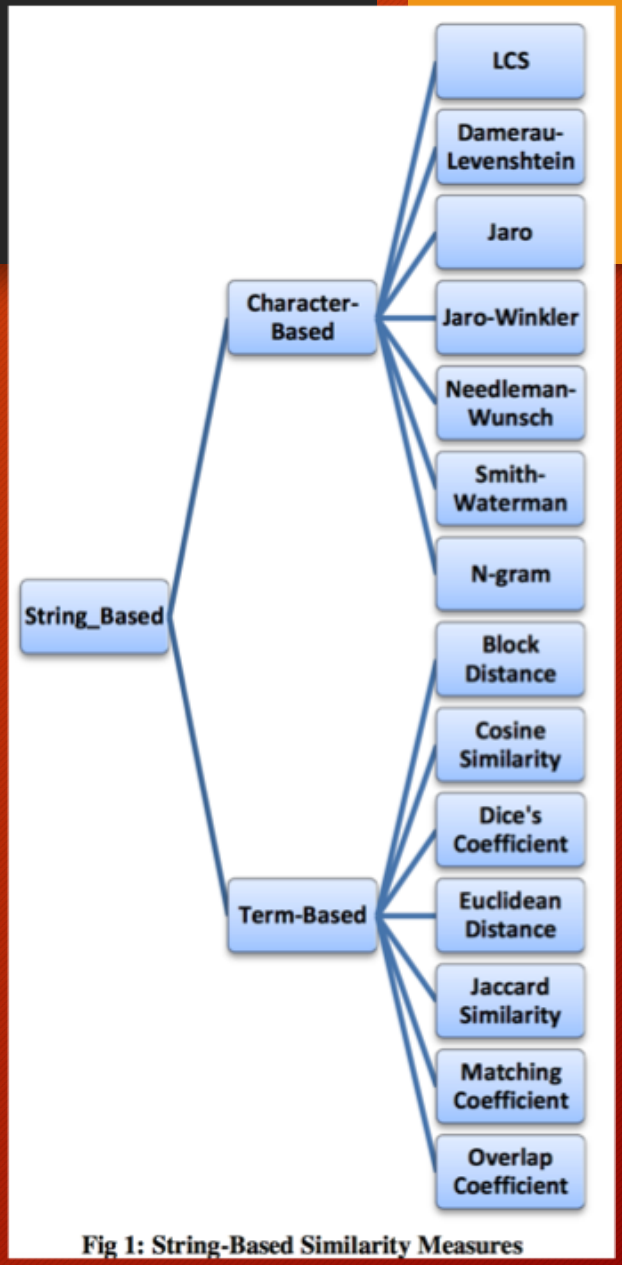


Fig 1: String-Based Similarity Measures

W. Gomaa, A. Fahmy. (2013). "A Survey of Text Similarity Approaches". Intl Journal of Computer Applications.
https://pdfs.semanticscholar.org/5b5c/a878c534aee3882a038ef9e82f46e102131b.pdf

Defender's Task

- Given some examples of what's malicious, find similar malicious examples in your own environment (if they exist)

Events in the millions, billions, more (?) per day

Relatively small volume of true malicious events

Which set of tools to use?

- Needle in a haystack: given a handful of examples

- Needle-stack of needles: I have thousands or more examples

Given a few needles, find a needle in a haystack

- Given a handful of examples, find similar examples in another environment
- String similarity
  - Levenshtein
  - Jaro
- Distance / similarity measurements
  - Cosine
  - Euclidean

String Distance / Similarity Measurements

- Levenshtein Distance (“Edit Distance”)
  - String metric representing the minimum number of edits e.g. insertions, deletions, and substitutions required to turn one string into another.

- Jaro Similarity
  - Measures several aspects of string similarity:
    - Number of matching characters
    - String length
    - Sequence of characters and transpositions

Wikipedia. (2018). https://en.wikipedia.org/wiki/Levenshtein_distance

String Distance / Similarity Measurements

String1	String2	Levenshtein	Jaro	Note
google.com	google.com	0	1.0	Exact match
google.com	g00gle.com	2	0.867	Replace letter o with number zero
google.com	google.badguy.com	7	0.863	Subdomain mimicking legitimate registered domain
google.com	malware.ru	9	0.533	Completely unrelated

Wikipedia. (2018). https://en.wikipedia.org/wiki/Levenshtein_distance

Wikipedia. (2018). https://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance

String Distance / Similarity Measurements

String1	String2	Levenshtein	Jaro	Note
bit	bot	1	0.778	Small change in short string
bitly.bit	bitly.bot	1	0.926	Small change in long string

Text to Numeric Vector

- Many techniques only work with numbers as input, rather than text
- This requires transforming words into numbers

Word Counts

badsite.pl/index.php?pl

URL #1	
Term	Term Count
badsite	1
index	1
pl	2
php	1

words in this sentence: 5

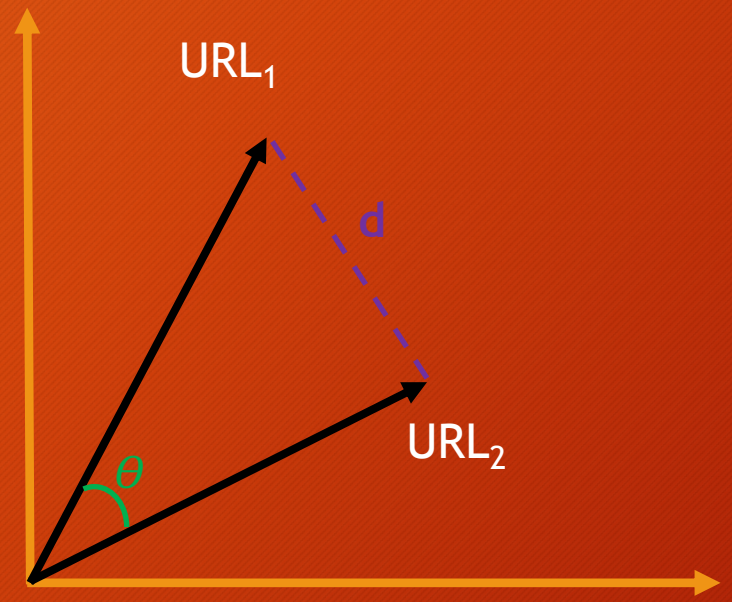
badsite.br/index?
click=br&br=click

URL #2	
Term	Term Count
badsite	1
index	1
click	2
br	3

term in this URL: 7

String Distance Measurements

- Text > Numbers > **Cosine Similarity** or **Euclidean Distance**



Term Frequency-Inverse Document Frequency

badsite.pl/index.php?pl

URL #1	
Term	Term Count
badsite	1
index	1
pl	2
php	1

words in this sentence: 5

badsite.br/index?
click=br&br=click

URL #2	
Term	Term Count
badsite	1
index	1
click	2
br	3

term in this URL: 7

Term Frequency-Inverse Document Frequency

badsite.pl/index.php?pl

URL #1	
Term	**Term Count**
badsite	1
index	1
pl	2
php	1

words in this sentence: 5

URL #1	
Term	**TF**
badsite	1/5 = .2
index	1/5 = .2
pl	2/5 = .4
php	1/5 = .2

badsite.br/index?
click=br&br=click

URL #2	
Term	**Term Count**
badsite	1
index	1
click	2
br	3

term in this URL: 7

URL #2	
Term	**TF**
badsite	1/7 = 0.14
index	1/7 = 0.14
click	2/7 = .29
br	3/7 = .43

Term Frequency-Inverse Document Frequency

badsite.pl/index.php?pl

URL #1	
Term	**Term Count**
badsite	1
index	1
pl	2
php	1

words in this sentence: 5

URL #1		
Term	**TF**	**IDF**
badsite	1/5 = .2	log (2 / 2) = 0
index	1/5 = .2	log (2 / 2) = 0
pl	2/5 = .4	log (2 / 1) = 0.301
php	1/5 = .2	log (2 / 1) = 0.301

IDF = $\log_e$(# docs / # docs with word)

badsite.br/index?
click=br&br=click

URL #2	
Term	**Term Count**
badsite	1
index	1
click	2
br	3

term in this URL: 7

URL #2		
Term	**TF**	**IDF**
badsite	1/7 = 0.14	log (2 / 2) = 0
index	1/7 = 0.14	log (2 / 2) = 0
click	2/7 = .29	log (2 / 1) = 0.301
br	3/7 = .43	log (2 / 1) = 0.301

Term Frequency-Inverse Document Frequency

badsite.pl/index.php?pl

URL #1	
Term	Term Count
badsite	1
index	1
pl	2
php	1

words in this sentence: 5

URL #1			
Term	TF	IDF	TF*IDF
badsite	1/5 = .2	log (2 / 2) = 0	0
index	1/5 = .2	log (2 / 2) = 0	0
pl	2/5 = .4	log (2 / 1) = 0.301	0.1204
php	1/5 = .2	log (2 / 1) = 0.301	0.0602

IDF = $\log_e$(# docs / # docs with word)

badsite.br/index?
click=br&br=click

URL #2	
Term	Term Count
badsite	1
index	1
click	2
br	3

term in this URL: 7

URL #2			
Term	TF	IDF	TF*IDF
badsite	1/7 = 0.14	log (2 / 2) = 0	0
index	1/7 = 0.14	log (2 / 2) = 0	0
click	2/7 = .29	log (2 / 1) = 0.301	0.0873
br	3/7 = .43	log (2 / 1) = 0.301	0.1294

Term Frequency-Inverse Document Frequency

badsite.pl/index.php?pl

URL #1	
Term	Term Count
badsite	1
index	1
pl	2
php	1

words in this sentence: 5

URL #1			
Term	TF	IDF	TF*IDF
badsite	1/5 = .2	log (2 / 2) = 0	0
index	1/5 = .2	log (2 / 2) = 0	0
pl	2/5 = .4	log (2 / 1) = 0.301	0.1204
php	1/5 = .2	log (2 / 1) = 0.301	0.0602

IDF = $\log_e$(# docs / # docs with word)

badsite.br/index?
click=br&br=click

URL #2	
Term	Term Count
badsite	1
index	1
click	2
br	3

term in this URL: 7

URL #2			
Term	TF	IDF	TF*IDF
badsite	1/7 = 0.14	log (2 / 2) = 0	0
index	1/7 = 0.14	log (2 / 2) = 0	0
click	2/7 = .29	log (2 / 1) = 0.301	0.0873
br	3/7 = .43	log (2 / 1) = 0.301	0.1294

Text to Numeric Vector

badsite.pl/index.php?pl

URL #1	
Term	**TF*IDF**
badsite	0
index	0
pl	0.1204
php	0.0602

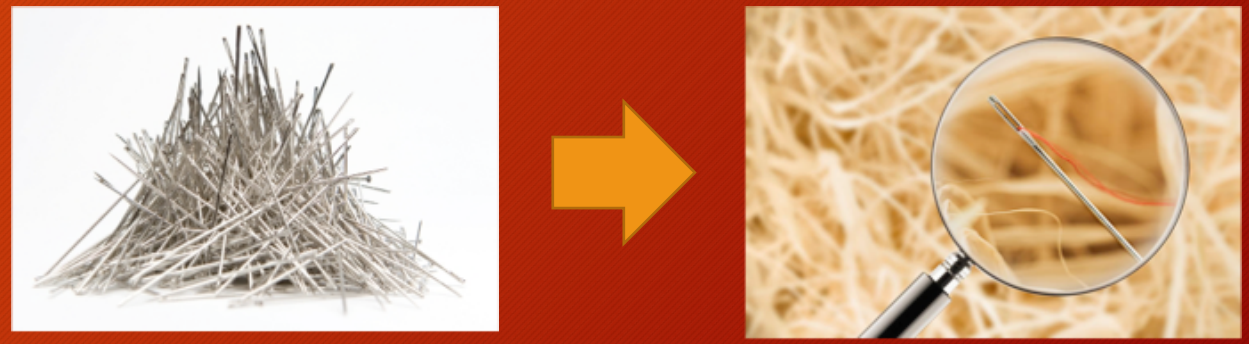
badsite.br/index?
click=br&br=click

URL #2	
Term	**TF*IDF**
badsite	0
index	0
click	0.0873
br	0.1294

URL	badsite	index	pl	php	click	br
#1	0	0	0.1204	0.0602	0	0
#2	0	0	0	0	0.0873	0.1294

Given a needle stack of needles, find a needle in a haystack

- Given many examples, find similar examples in another environment
- String similarity
  - Naïve Bayes
  - Logistic Regression
  - Decision Trees
  - Random Forest
  - Neural Networks
  - and more…



Naïve Bayes Example

Corpus

1. I loved the movie
2. I hated the movie
3. A great movie. A good movie.
4. Poor acting
5. Great acting, a good movie

Doc	I	loved	the	movie	hated	a	great	poor	acting	good	label
#1	1	1	1	1							+
#2	1		1	1	1						-
#3				2		1	1			1	+
#4								1	1		-
#5				1		1	1		1	1	+

Example for word "good":

$P(\text{"good"}|\ -) = (N_{k-} + 1) / (n_- + |\text{vocab}|) = 0.0625$

$P(\text{"good"}|\ +) = (N_{k+} + 1) / (n_+ + |\text{vocab}|) = 0.125$

Example: What's the value of "I hated the poor acting"?

$V(+) = P(+)P(I|+)P(\text{hated}|+)P(\text{the}|+)P(\text{poor}|+)P(\text{acting}|+) = 6.05*10^{-7}$

$V(-) = P(-)\ P(I|-)P(\text{hated}|-)P(\text{the}|-)P(\text{poor}|-)P(\text{acting}|-) = 1.22*10^{-5}$

Francisco Iacobelli. (2015). Text Classification Using Naïve Bayes https://www.youtube.com/watch?v=EGKeC2S44Rs

Questions

Naïve Bayes Example

Given these definitions…

N_{k+} = # times word K occurs in the + class

N_{k-} = # times word K occurs in the - class

|vocab| = # unique words in vocabulary = 10

$P_+ = 3 / 5 = .6$

$P_- = 2 / 5 = .4$

n_+ = sum of word counts for that class = 14

n_- = sum of word counts for that class = 6

W_k = a word, like "loved"

…we get the probability of any word for it's label

$P(W_k \mid +) = (N_{k+} + 1) / (n_+ + |vocab|)$

$P(W_k \mid -) = (N_{k-} + 1) / (n_- + |vocab|)$

Example for word "good":

$P(\text{"good"} \mid -) = (N_{k-} + 1) / (n_- + |vocab|)$

$= 0 + 1 / 6 + 10 = 1 / 16 = 0.0625$

$P(\text{"good"} \mid +) = (N_{k+} + 1) / (n_+ + |vocab|)$

$= (2 + 1) / (14 + 10) = 3 / 24 = 0.125$

Example: What's the value of "I hated the poor acting"?

$V(+) = P(+)P(I|+)P(hated|+)P(the|+)P(poor|+)P(acting|+) = 6.05*10^{-7}$

$V(-) = P(-)\ P(I|-)P(hated|-)P(the|-)P(poor|-)P(acting|-) = 1.22*10^{-5}$

*Based on the Naïve Bayes model the sentence is more likely to be negative (-), since $1.22 * 10^{-5}$ is bigger than $6.05*10^{-7}$.