Human Factors in Cybersecurity, Vol. 53, 2022, 1–8
https://doi.org/10.54941/ahfe1002195

A Metric to Assist in Detecting International Phishing or Ransomware Attacks

Wayne Patterson[1] and Jeremy Blackstone[2]

[1] Patterson and Associates, Washington, DC 20002, USA
[2] Howard University, Washington, DC 20059, USA

ABSTRACT

Over the past decade, the number of cyberattacks such as ransomware, phishing, and other forms of malware have increased significantly. The ability to launch such devastating attacks is no longer limited to highly structured organizations including government agencies whose missions may well include cyberattacks. The focus of our study is on threats to an individual not from such organizations, but rather less organized cybercriminal groups with limited resources. The Internet provides ample opportunities for such criminal organizations to launch cyberattacks at minimal cost. One tool for such lower-level criminal organizations is Google Translate (GT) needed to launch a cyberattack on a user in a relatively advantaged country such as the United States, United Kingdom, or Canada. It or has been observed that many such attacks may originate in a lesser developed country (LDC), where the local language is a language not common persons in target countries, for example English.

Keywords: Phishing, Ransomware, Cyberattack, Language, Levenshtein distance

DETECTING SUSPECTED TRANSLATIONS OF CYBERATTACK TECHNIQUES

It is a reasonable assumption that informal cyberattackers may not have a command of English and to use English for an attack online they may require a mechanism, such as the no-cost GT.

In previous work, a number of authors have attempted to develop an index to measure the efficiency or what might be called an ABA translation. This involves beginning with a test document in language A, then GT to translate into language B, then back again to A. The resulting original text is then compared to the transformation by using a modified Levenshtein distance computation for the A versions.

The paper analyzes the process of determining an index to detect if a text has been translated from an original language and location, assuming the attack document has been written in one language and translated using GT into the language of the person attacked. The steps involved in this analysis include:

© 2022. Published by AHFE Open Access. All rights reserved.

a) Consistency: in order to determine consistency in the use of the ABA/GT process, the primary selection of test is compared with random samples from the test media;
b) Expanded selection of languages for translation: prior work has established use of the technique for 12 language pairs (Patterson, 2022), (Florea and Patterson, 2021). The current work extends analysis to a wider set of languages, including those reported as having the highest levels of cyberattacks. This component is also addressed in greater detail in (Blackstone and Patterson, 2022).
c) Back translation of selected languages: used to extend the quality of those translations are made.
d) New language pairs are considered: by analyzing the countries and indigenous languages of the countries paired with the highest levels of cyberattack and the highest levels of cyberdefense, additional language pairs are added to this analysis;
e) Comparison to prior results: results found in this paper are used for a proposed network for all language pairs considered in this analysis.

The end product is a metric giving a probability of determining the original source language of the cyberattack as compared to the translation to the victim's language, with the expectation that this will allow for an increased likelihood of being able to identify the attackers.

INTRODUCTION

Objective

The purpose of this paper is to propose a series of techniques that can provide an approach to determining if a malicious cyberattack, such as ransomware or phishing, can be determined as to the attack language or country of origin. Many attacks of the types described above, in communicating with the target, must transmit text in some format. For example, in a ransomware attack, the target must be told that certain steps are necessary for the ransomware to be paid. Similarly, a phishing attack will also normally try to trap the target into a response that will provide a benefit to the attacker.

In most cases, cyberattacks of this nature must provide some information as text in a human language, in order to describe to the potential victim the actions that the attacker expects to be carried out. Since there are literally hundreds of natural languages throughout the world, it is entirely possible and indeed likely that the cyberattacker will need to provide some text in the language of the persons being targeted. Given that the potential attacker can come from virtually any corner of the world, it will often be the case that the cyberattackers command of his or her language will not include the language of the target. Thus, it will be necessary that any text information will need to be first translated from the human language of the attacker to a human language presumed to be understood by the victim.

With the existence of the Internet, we must assume that any cyberattack may originate in any part of the world, and thus also that the attacker needs to have a mechanism for his or her expected target to be able to read and understand the attack text in their own language, in order for the attacker to present the challenge to the potential victim, and thus potentially reap the reward. Given the large number of human languages in use throughout the world, if we wish to narrow down the potential source over a attack, it would be useful if we could examine the attack text and try to determine if had been translated between two human languages. In such a case, a fairly definitive conclusion could isolate the potential source of the attack.

Potential Determination of Cyberattack Source

Given the number of human languages in use throughout the world, it would be literally impossible to determine the exact physical source of a translated cyberattack. However, by studying the preponderance of usage of many of the world's more widely used languages, it should be possible, when detecting the source language of a translated cyberattack, to narrow down the origin of that attack.

We also make the assumption that most cyberattackers that focus on ordinary users do not have the vast resources available to, for example, a national government or a large corporate entity. Assuming this to be the case, the cyberattacker is not likely to have the resources to be able to easily produce an attack in the language most familiar to the potential victim. As a consequence, we make the assumption that the attacker, needing to produce the text of an attack in a language understandable to the potential victim, and also without vast resources to the highest quality of language translation software, will use the most easily available, and most widespread, and at no cost software, which for many users throughout the world, would be Google Translate.

Languages and Countries Considered

This study will analyze text assumed to have been produced in the principal languages of 20 different countries. The particular choice of countries follows several paths: 1) several countries were selected based on prior research results reported; 2) several were chosen as a result of Internet hacking levels; and 3) several were chosen for a geographic dispersion throughout the world. the following table demonstrates the rationale for each concrete choice.

In order to provide a potential tool in determining the original language source of a malicious cyberattack, we will begin with two tools and approaches to develop a database to assist in determining the language of a suspected translated cyberattack.

ABA Translation

In order to develop an approach to determine if a given body of text had been originally translated between human languages, we developed a technique called ABA translation. Using English as a base for our various analyses, we developed a body of English text from two categories: classical quotations (Q) that are well known in English, and that demonstrate grammatically correct use of the English language; and a category of popular expressions from more colloquial English language, in particular popular film (F). In previous cases, we had developed a selection of 20 texts, ten each from categories

Table 1. Countries and languages under consideration.

Country	Population (M)	Primary Language	Approx # Speakers	Minority Language	Approx # Speakers
Albania	2,872,933	Albanian	2,838,458	Greek	14,365
Brazil	213,993,437	Portuguese	213,993,437	-	-
Bulgaria	6,896,663	Bulgarian	5,296,637	Turkish, Romani	827,600
China	1,444,216,107	Mandarin	1,444,216,107	-	-
Estonia	1,325,185	Estonian	910,402	Russian	392,255
Ethiopia	117,876,227	Oromo, Amharic	74,026,271	Somalian	7,308,326
Finland	5,548,360	Finnish	4,821,525	Swedish	288,515
France	65,426,179	French	65,426,179		-
Germany	83,900,473	German	83,900,473		-
Greece	10,370,744	Greek	10,267,037		-
India	1,393,409,038	Hindi	635,394,521	Bengali	111,472,723
Italy	60,367,477	Italian	60,367,477		-
Latvia	1,866,942	Latvian	1,161,238	Russian	631,026
Lithuania	2,689,862	Lithuanian	2,262,174	Russian	215,189
North Macedonia	2,082,658	Macedonian	1,384,968	Albanian	522,747
Romania	19,127,774	Romanian	16,335,119	Hungarian	1,166,794
Russia	145,912,025	Russian	125,046,605		-
Saudi Arabia	35,340,683	Arabic	35,340,683		-
Serbia	8,697,550	Serbian	7,662,542		-
Spain	46,745,216	Spanish	34,591,460	Castillian	7,946,687

Q and F. For consistency we have used the same quotes in several previous publications. The ABA approach chooses a text in English, translates it into the language we wish to study, and then back to English, hence ABA.

Levenshtein Distance

In order to determine a metric for the quality of the translation based on the text samples, we use a well-known approach called the Levenshtein Distance to determine the accuracy of the ABA translation. Given this approach we can determine a metric with the assumptions above, for the translation of text between any of the given language pairs. For our purposes, we use a version of the standard definition that we designate as Modified Levenshtein Distance (MLD). In this case, we break up the computation of this distance for substrings of all the strings being compared. Thus the opening few characters of the strings being compared will not propagate through the entire set of strings. Here are a few examples to demonstrate computing MLD.

Estonian ([1] Estonian --- [2] Original English --- [3] Translation Back to English

[1] Ma olen sama vihane kui põrgu ja ma ei võta seda enam vastu!
[2] I'm as /mad /as hell, and I'm not /going to /tak /e this /anymore!
[3] I'm as /angry /as hell, and I'm not / /tak /ing it /anymore!
/ 5/ /7/ / 5/→ MLD = 5 + 7 + 5 = 17

Table 2. Level of ABA Errors in Test Languages.

Country	Sum of Misses	% of Errors	Country	Sum of Misses	% of Errors
Albania	260	26.50%	India	387	39.45%
Brazil	185	18.86%	Italy	208	21.20%
Bulgaria	191	19.47%	Latvia	236	24.06%
China	453	46.18%	Lithuania	284	28.95%
Estonia	252	25.69%	N.Macedonia	237	24.16%
Ethiopia	526	53.62%	Romania	211	21.51%
Finland	297	30.28%	Russia	420	42.81%
France	193	19.67%	Saudi Arabia	324	33.03%
Germany	267	27.22%	Serbia	273	27.83%
Greece	280	28.54%	Spain	215	21.92%

Bulgarian ([1] Bulgarian - [2] Original English - [3] Translation Back to English

[1] От всички джин стави във всички градове по целия свят, тя влиза в моя.

[2] Of all the gin joints in all /the towns in all /the world, /she walks into/mine.

[3] Of all the gin joints in all /cities around /the world,/it enters /mine.

/ 13/ / 12/ → MLD=13+12=25

German ([1] German --- [2] Original English --- [3] Translation Back to English

[1] Ich lebe so weit über mein Einkommen hinaus, dass man fast sagen kann, wir leben getrennt.

[2] I /'m living /so far beyond my income that /we may /almost /be said to be /liv/ing apart.

[3] I / live /so far beyond my income that /you can /almost /say we /liv/e separately.

/ 8/ / 6/ / 8/ /
11 → MLD = 8 + 6 + 8 + 11 = 33

Baseline Assessments of MLD for All Test Languages

The first step in developing an index for identifying the language of origin of The GT translations is to compute the accuracy for the tests for the 20 test samples in the 20 selected languages. The following table demonstrates the Levenshtein Distance results for these 20 uses of GT in the chosen languages.

This identification of differences in translation with various language sets is a useful beginning for this study. However, using only a single measure will not assist greatly in differentiating between several of the languages studied in this paper. For example, text translated using the ABA approach will show the difference between translating into Bulgarian versus translating into French yields only about a 0.2% difference; whereas taking three countries and languages into consideration (Italy, Romanian, and Spain) again shows only about a 0.3% difference between the first two and a 0.4% difference between the latter two.

Instead, we use a series of five measures to give a combined differentiation between the languages studying, and these results are far more helpful.

Table 3. Countries and languages under consideration.

Country	m	f	u	w	l	Country	m	f	u	w	l
Albania	22.0	24.8	55.6	67.7	7.5	India	59.2	69.2	33.3	20.0	93.1
Brazil	0.0	0.0	88.9	69.8	72.8	Italy	6.7	24.1	55.6	56.1	55.6
Bulgaria	1.8	6.8	66.7	65.0	20.9	Latvia	15.0	13.5	77.8	88.7	4.8
China	78.6	100.0	0.0	65.6	100.0	Lithuania	29.0	31.6	33.3	81.3	9.0
Estonia	19.6	27.1	55.6	88.9	0.0	N. Macedonia	15.2	31.6	66.7	79.2	5.2
Ethiopia	100.0	89.5	11.1	0.0	44.5	Romania	7.6	7.5	33.3	75.5	38.0
Finland	32.8	46.6	33.3	92.8	21.9	Russia	68.9	84.2	22.2	83.8	69.0
France	2.3	5.3	100.0	81.7	58.1	Saudi Arabia	40.8	51.9	22.2	100.0	49.6
Germany	24.0	36.1	44.4	89.8	62.3	Serbia	25.8	15.8	44.4	75.4	25.9
Greece	27.9	21.8	33.3	75.1	28.9	Spain	8.8	12.0	44.4	94.1	53.1

Consider the computation of the following statistic for each of the 20 languages and countries being considered, which we call COUNTRIES. We refer to individual countries as COUNTRY(i), i= 1,...,20. Furthermore, we define five "coordinates" for each country, say X ∈ { m, f, u, w, l }, and the coordinate values are "normalized", distributing each in a range from 0 to 100, by computing:

$$X_{n,i} \;=\; 100 \times (\; X_i - \min(X_i))/(\max(X_i - \min(X_i)\;)\;)$$

The definitions are: m (misses in MLD for all 20 samples); f (misses in MLD for only the 10 F film samples); u (number of text translation samples with < 10 errors); w (percent of web users of country population); and l (logarithm of the country population). Let x, y ∈ COUNTRIES, x ≠ y. Then compute $d_{x,y}$ for all x, y as follows:

By using all five "coordinates", the values of the $d_{m,n}$ as described above will give both the minimum distances between the individual values for each pair of countries (380 pairs). The minimum distance among these 380 pairs is 168.36 (between Greek and Romanian). In addition to having distinct values for each country under consideration, we can also find the minimum distance between any pair of country values. Then, should some new set of values for an unknown country be tabulated, as long as the distance between the new and any of the base countries is less than 1/2 the minimum distance between any pair of countries, e.g. 84.18, it is a reasonable assumption to make that assuming the new country is one of the existing ones, that's a set of values can be associated with newest country in the computation. If we have a set of values { m, f, u, w, l } for some country or language Y, as long as the $d_{x,n}$ < 84.18 = 168.36/2, we can use the minimum distance principle to associate the values to the nearest country dataset.

Consistency

In order to validate the data developed, it is needed to verify that there is consistency throughout all of the data points used. In other words, by taking a random subset of the 20 translations, we compare the relationships of the subset to the relationship to the whole to try to verify the consistency of

Table 4. ABA and BAB comparison.

Country	ABA	BAB	% DIFF	Country	ABA	BAB	% DIFF
Albania	260	142	45%	India	387	171	56%
Brazil	185	157	15%	Italy	208	143	31%
Bulgaria	191	261	37%	Latvia	236	186	21%
China	453	105	77%	Lithuania	284	263	7%
Estonia	252	267	6%	N. Macedonia	237	265	12%
Ethiopia	526	298	43%	Romania	211	250	18%
Finland	297	276	7%	Russia	420	208	50%
France	193	150	22%	Saudi Arabia	325	108	67%
Germany	267	172	36%	Serbia	273	193	29%
Greece	280	160	43%	Spain	215	84	61%

the translated data. Using this approach, the differences between the full set of data points and a random subset of the data points are less than 5% in 17 of 20 cases. The validity of the GT algorithm can be approximated by a comparison of the forward (ABA) and back translations (BAB).

As mentioned above, (Blackstone and Patterson, 2022) establishes a more thorough analysis of a sample of translation of actual cyberattacks. On our limited example of these, we see the translation process has very comparable accuracy to our test base. The test malware examples can be found in (Patterson and Winston-Proctor, 2019, pp. 32, 36, 45). With respect to the comparative effectiveness of the translations, the translations of malware examples have fewer MLD values in 65% of the cases, and the differences of the MLD translations are less than 30% in 16 of the 20 comparisons.

CONCLUSION

This research has been done with a relatively small subset of all countries and major languages used. Thus the results reported here provide a starting point for a similar analysis for determining cyberattacks that use text in the attack itself, such as is essentially necessary in ransomware or phishing attacks. Further study will include (a) more extensive data sources for translation analysis; (b) an enlarged set of countries and languages; and (c) alternate choices for the "vectors" to be in measuring distances between pairs of countries and languages.

REFERENCES

Blackstone, J. Patterson, W. (2022). Isolating Key Phrases to Identify Ransomware Attackers. Proceedings of the 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022), New York, NY.

Florea, D. Patterson, W. (2021). A Linguistic Analysis Metric in Detecting Ransomware Cyber-Attacks, www.thesai.org

Patterson, W. and Winston-Proctor, C. (2019). *Behavioral Cybersecurity*. CRC Press.

Patterson, W. Murray, A. Fleming, L. (2020). Distinguishing a Human or Machine Cyberattacker. Proceedings of the 3rd Annual Conference on Intelligent Human Systems Integration, Modena, Italy, February 2020, pp. 335–340.

Patterson, W. (2022). Detecting Cyberattacks Using Linguistic Analysis. Proceedings of the 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022), New York, NY.

WPR (2021). Top 20 Countries Found to Have the Most Cybercrime. World Population Review. https://worldpopulationreview.com/countries