International Journal of Cyber Criminology
Vol 14 Issue 1 January – June 2020

What's in a Name? Using Words' Uniqueness to Identify Hackers in Brute Force Attacks

Amit Rechavi[1] & Tamar Berenblum[2]
Hebrew University of Jerusalem, Jerusalem, Israel

Abstract

Do hacker subgroups share unique practices and knowledge? Is there a spatial characteristic to this sharing? The study investigates whether hackers who perform brute force attacks (BFAs) from different countries (different IPs) use a spatially based corpus of words for usernames and passwords. The study explores the usage of 975,000 usernames (UNs) and passwords (PWs) in brute force attacks on honeypot (HP) computers. The results suggest that hacker subgroups attacking from different countries use different combinations of UNs and PWs, while a few attacks coming from different IPs share the same corpus of words. This significant result can help in tracing the source of BFAs by identifying and analyzing the terms used in such attacks.

Keywords: Hackers, SNA, Brute Force Attacks (BFAs), Honeypot, Knowledge Exchange.

Introduction

Hacking is the act of identifying weaknesses in information technology systems or networks, exploiting their vulnerabilities to gain access, and executing fraudulent actions such as fraud, privacy invasion, and the stealing of corporate/personal data. Hacking can be performed in a variety of ways, one of which is a brute force attack (BFA). In such an attack, hackers try to identify the needed information to access a device—both usernames (UNs) and passwords (PWs). These attacks entail the need to run various combinations of UNs and PWs until the hacker gains access to the device.

While hacking is not a new topic in cyber criminology, much of the literature is based on off-line qualitative data (such as interviews) and focuses on hacker subculture and motivations (Jordan & Taylor,1998; Barber, 2001; Turgeman-Goldschmidt, 2005).

In the last decade, there has been growing interest in understanding the hacking modus operandi, trust-building among this community's members (Von Lampe, & Ole, 2004; Leukfeldt et al., 2013; Dupont et al., 2016), and hacker co-offending behaviors (Holt, 2007; Holt et al., 2012; Malm et al., 2011; Rechavi et al., 2015).

[1] Business Administration, Ruppin Academic Center, Natanya, Israel and The Federmann Cyber Security Center, Hebrew University of Jerusalem, Jerusalem, Israel.
Email: amit.rechavi@gmail.com
[2] The Federmann Cyber Security Research Center, Hebrew University of Jerusalem, Jerusalem, Israel. Email: tamar.berenblum@mail.huji.ac.il

As part of this trend, there is a growing interest in understanding the social network of hackers. Social network analysis (SNA) was established in the field of criminology (Bouchard & Amirault, 2013) and is regarded as useful in providing measures of the structure of criminal networks and the position of key players in those networks, as well as the understanding of the effect of the removal of nodes from these illicit networks. (See, for example, Bright et al., 2017; Décary-Hétu & Dupont, 2012; and Easton & Karaivanov, 2009).

The literature indicates that hackers form a global community consisting of micro-communities that exist in the on-line and off-line worlds (Holt, 2007; Moilanen, 2012). These communities are autonomous and dynamic (Moilanen, 2012) and differ from each other concerning the type of activities they choose, the technologies they use, and the level of off-line connections they have. For example, hackers from Europe and North America participate in 'real-world communities' and 'hackerspaces' more than hackers from Asia, Australia, and South America do (Moilanen, 2012).

Hacker communities are, in essence, learning communities (Zhang et al.,2015), and their members share information and knowledge in many channels (Holt, 2007; Olson, 2012; Summers, 2015; Rechavi et al., 2015), including public online forums, newsgroups, web sites, and subcultural publications (Kleinknecht, 2003, Chavez & Bichler, 2019) as well as conventions (Summers, 2015). Such information dissemination and even the spread of gossip (Chavez & Bichler, 2019) among the online hacker community enable its members to learn skills, build their prestige, explore problems and reflect on possible solutions (Summers, 2013) and learn the norms and values of the community as well as how to make sense of and justify their actions (Holt, 2005; Holt & Copes, 2010).

This culture of information sharing has characterized the hacking community from its early days (Thomas, 2002) and shifted throughout the years. In the first generation of hacking, sharing information was the everyday norm, while in the third generation, by which time hacking had become associated with criminality (Chandler,1996), the community became more competitive and secretive about sharing information methods and scope.

By examining knowledge transfer and sharing, Holt (2013) finds that there are virtual markets for stolen data, which make economic crimes much more comfortable to commit, and these buying and selling processes are peer-driven. The users of these markets practically divide the trading efforts between the participants. There are four types of hackers, differentiated based on their knowledge transfer characteristics (Zhang et al., 2015), and there are hackers who produce knowledge and others who consume it (Holt & Kilger, 2012). This phenomenon is well-known in social networks where some members explore data and some exploit it (Lazer & Friedman, 2007).

Understanding the modus operandi of the hacker community and the information sharing facilitated by its networks is a crucial step in identifying a hacking network and its members in an attempt to prevent that network's harmful actions. This study focuses on the prevention of hacking performed by using brute force attacks (BFAs). The literature on the prevention of BFAs (Pliam, 2000; Sullivan, 2007; Weir et al.,2009; Raza et al., 2012; Dave, 2013) and the right way to choose UNs and PWs as preventive mechanisms (Bonneau, 2012; Kelley et al., 2012; Juels & Rivest, 2013; Egelman et al., 2013; Mazurek et al., 2013; Zhang-Kennedy et al., 2013; Ma et al., 2014; Jose et al., 2016; Gagneja & Jaimes,2017) are based on studying successful hacking attempts. None of the studies we

are aware of include hackers' failed attempts or their data. Moreover, these studies focus on technological solutions.

Thus, in this study, we aim to contribute to the literature on the prevention of BFAs by exploring the data that hackers use in their penetration attempts—those that are successful as well as those that fail. The vast amount of data allows us to better understand hacking patterns and information sharing. Using data on such efforts, we try to identify whether hackers cooperate and, despite being a heterogeneous community (Barber, 2001), share hacking techniques and tools.

Moreover, we suggest a new opportunity to address the need to identify the attackers' locations to disrupt the hacking network: exploring the linguistic characteristics of the data as a possible clue to hackers' locations. We suggest a way to identify the spatial source of BFAs by analyzing the words (passwords and usernames) that are used by hackers. Differentiating hackers by analyzing their corpus of words will allow a better, more precise mapping of BFAs and the connections between hackers.

Research Question

To achieve these goals, the study aims to determine whether hackers (individuals or subgroups) use unique combinations of usernames and passwords. Do they share these combinations with others? Is there a spatial pattern of this sharing? What can one learn from this pattern about hackers' locations?

The next chapter presents the results in three sections:

(**A**) First, it shows the distribution of attacks for 35 days. An analysis of the distribution of the attacks raises the possibility that the algorithms executing these attacks differ from country to country.

(**B**) Then, the study presents the patterns among hackers' IP addresses, usernames, and passwords. The analysis raises the possibility that different IPs use different sets of usernames and passwords.

(**C**) Finally, the study presents the network of IPs who share pairs of usernames and passwords in BFAs and presents the main finding of the analysis of this network.

Data and Methods

In an attempt to explore the relations between hackers' profiles and the words they use to guess UNs and PWs in BFAs, the study collects and analyzes unsuccessful hacking attacks on honeypots (HPs), builds a bipartite network of "hackers" and "words" and analyzes that network (*see Appendix A*). The research includes the establishment of 157 HPs located on an academic campus in China and the recording of 975K hacking attempts over 35 days executed from 314 IPs in 54 different countries. The honeypots were Windows-based servers equipped with software that could identify the origin of the attacks and software that could record and collect the keystrokes of hackers trying to penetrate the servers.

After 35 days, the researchers had 975K records, each of which included details concerning the attacker and the attack.

The researchers removed all duplicated records and mapped those containing the origin of the attack, the username, and the password. (For example, the combination UN=root and PW=123 appears 200 times in attacks coming from an IP in China, but in our dataset, it appears only once.)

After cleaning the data, the final dataset includes 55 countries, approximately 1,500 usernames, and 12,700 passwords. For example, the dataset consists of a BFA from China with UN=root and PW=123 and a BFA from Israel with UN=root and PW=123. These are two different and unique records. (Appendix B details several statistics concerning leading IP attacks.) The study mapped the relationships between hackers' IPs, UNs, and PWs. Using unique pairs of UN and PW, the researchers created a network of connected countries.

The researchers are aware of the fact that though a BFA comes from a specific IP address, it might be initiated in other IPs. However, even the identification of these IPs (which are not necessarily the IPs that launched the attack) might help policymakers and companies alert, control, and regulate suspicious IPs.

Using the collected data (the UNs, PWs, and hackers' IPs), the study builds two bipartite networks. Each network includes two types of nodes: humans (hackers from different IP addresses) and the "words" they use (usernames, passwords). In reality, each IP might use many UNs and PWs, and each UN and PW might be used in attacks initiated by different IP addresses. These relations between hackers' IP addresses the words they use are defined as "many-to-many" relationships. These relations are the basis for modeling the attacks as networks. The researchers built the networks using ORA[3] software. One network models and maps the hacker IPs and the UNs they use, and the other network models and maps the hacker IPsand the PWs used.

The researchers analyzed the two networks with SNA methodology (see Benjamin & Chen, 2012; Lu et al., 2010; Rechavi et al., 2015) and explored three levels of analysis in each network:

(1) a single node's centralities (the volume of centrality of the specific hacker's IP, the specific user-name, and the specific password);

(2) the existence of groups and cliques in the networks and an analysis concerning their size, density, reciprocal connections, connections with each other and so on; and

(3) an analysis of the entire network (including parameters such as total number of nodes and links, density, diameter, and distribution of centrality characteristics).

Results

A. Distribution of attacks for 35 days

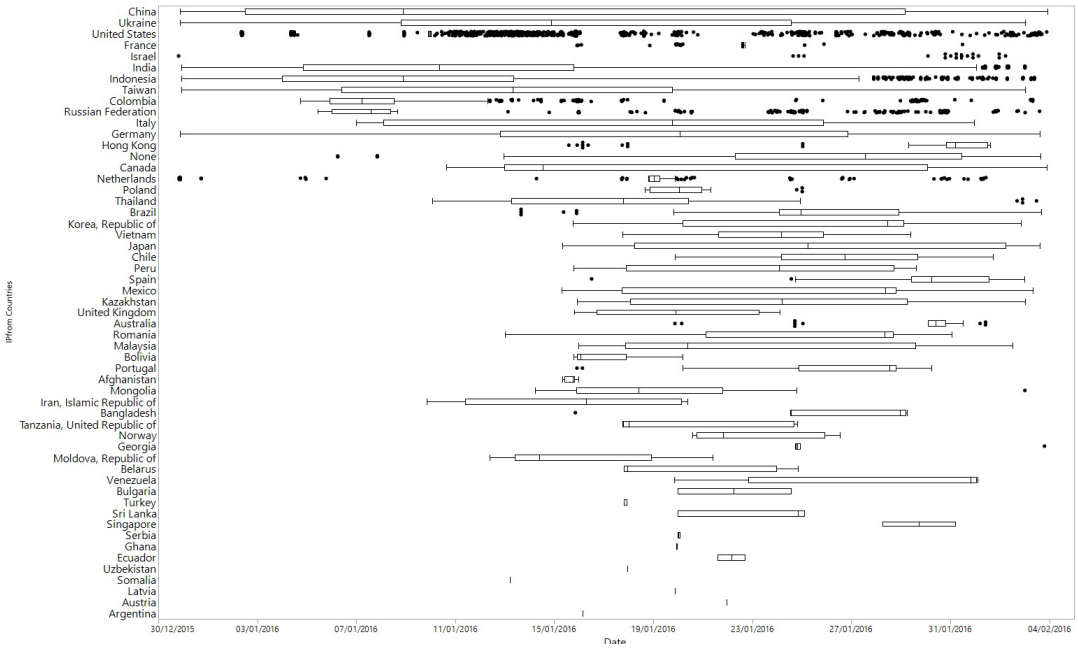
This study analyzes the patterns of the 975K attacks across 35 days. Graph 1 presents the distribution of the IP addresses of the attacks.

Graph 1 presents the fact that hackers (using IPs from different countries) execute different patterns of attacks. It is clear that the attacks do not occur on the same days and do not have a uniform distribution in all IPs. For example, attacks initiated from Canada and Germany occur mostly in the middle of the month, attacks launched from India and Indonesia occur primarily in the first days of the month, while attacks coming from US IPs are distributed across the entire month. The graph illustrates the different patterns of attack coming from different IPs, and it supports the assumption that different IPs use different methodologies (or tools, if the attacks are originated by automated software, known as bots).

[3] http://www.casos.cs.cmu.edu/projects/ora/

Next, the study explores the different patterns of attacks coming from different IPs on an hourly basis.

Graph 1. Distribution of the attacks by countries' IPs

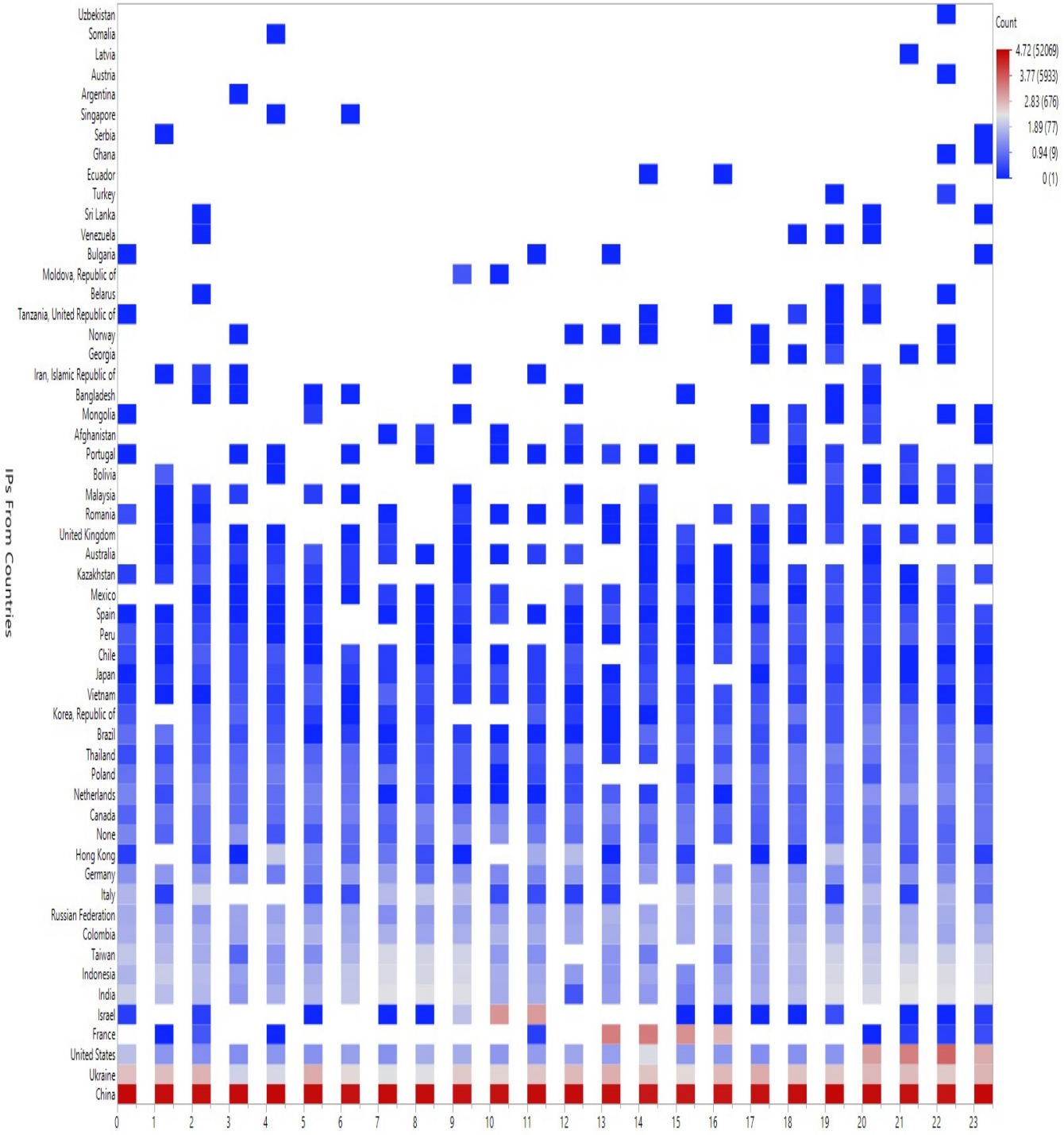


Graph 2 presents the hourly distribution of attacks coming from different IPs. The hours of the attacks present the local time at the origin of the attack (based on the IP location). As one can see from graph 2, the pattern of attacks during a given 24-hour period changes from country to country. Some IPs attack HPs continuously all day long, while others have specific hours for attacks.

Some of the attacks are low in number, sparsely distributed, and random; however, looking at the countries that initiate most of the BFAs (China, Ukraine, US, and France), clear patterns emerge. For instance, attacks from Ukraine and China are continuous attacks 24/7, while the primary US attacks occur between 20:00 and 23:00, and attacks coming from France occur between 13:00 and 16:00.

The different daily patterns of attacks and the different hourly patterns of attacks make it reasonable to assume that there are various algorithms involved in the BFAs coming from different countries. These findings call for additional analysis of these attacking algorithms. Next, the study explores how these algorithms use word dictionaries. In case the study will find that specific IPs use specific dictionaries (in the process of guessing usernames and passwords), the possibility that different algorithms are applied in various places in the world becomes even higher.

Graph 2. Hourly pattern of the attacks for 35 days, distribution by countries' IPs.
(The color of the box is following the log of its attack in the specific hour.
Dark red means more than 3,000 attacks).

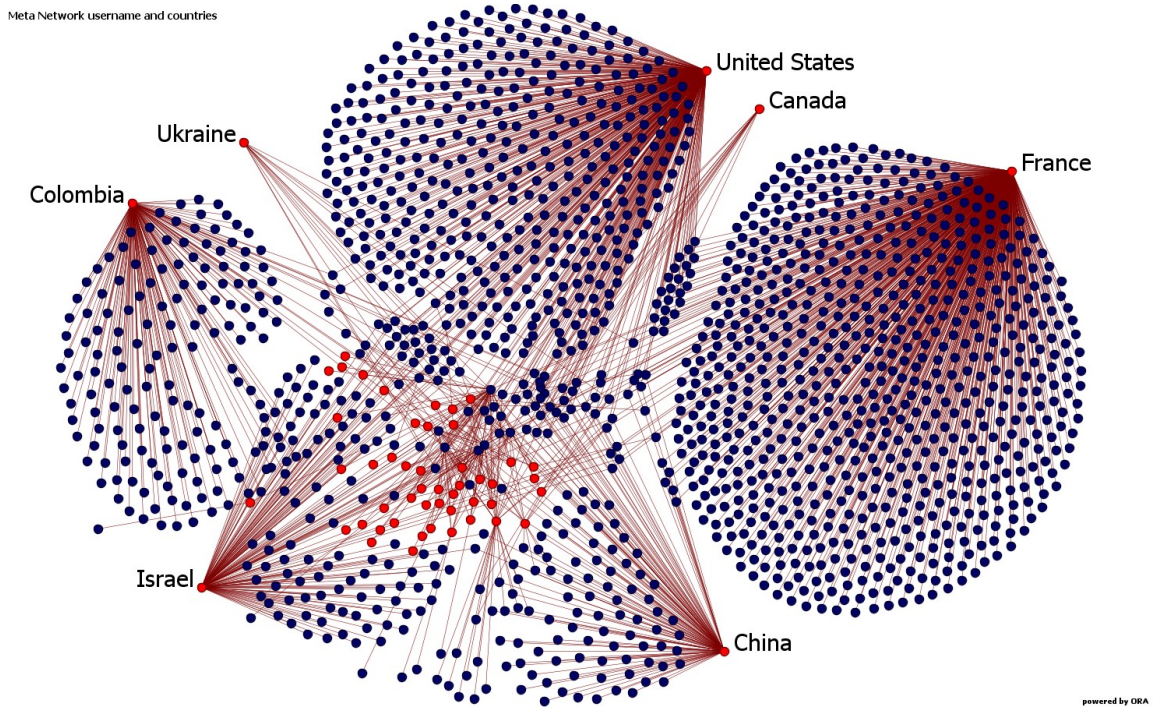


B. The network between IPs and PWs

The study maps the relations between hackers' IP addresses and (1) UNs and (2) PWs. Based on these relations, the study builds a network and applies SNA methodologies to explore this network. The two networks are bipartite (have two types of nodes), where IP addresses and UNs (in Graph 3) and IP addresses and PWs (in Graph 4) are the nodes.

Graph 3 presents the relationships between hackers' IPs and UNs, and Graph 4 presents the relationships between Hackers' IPs and PWs.

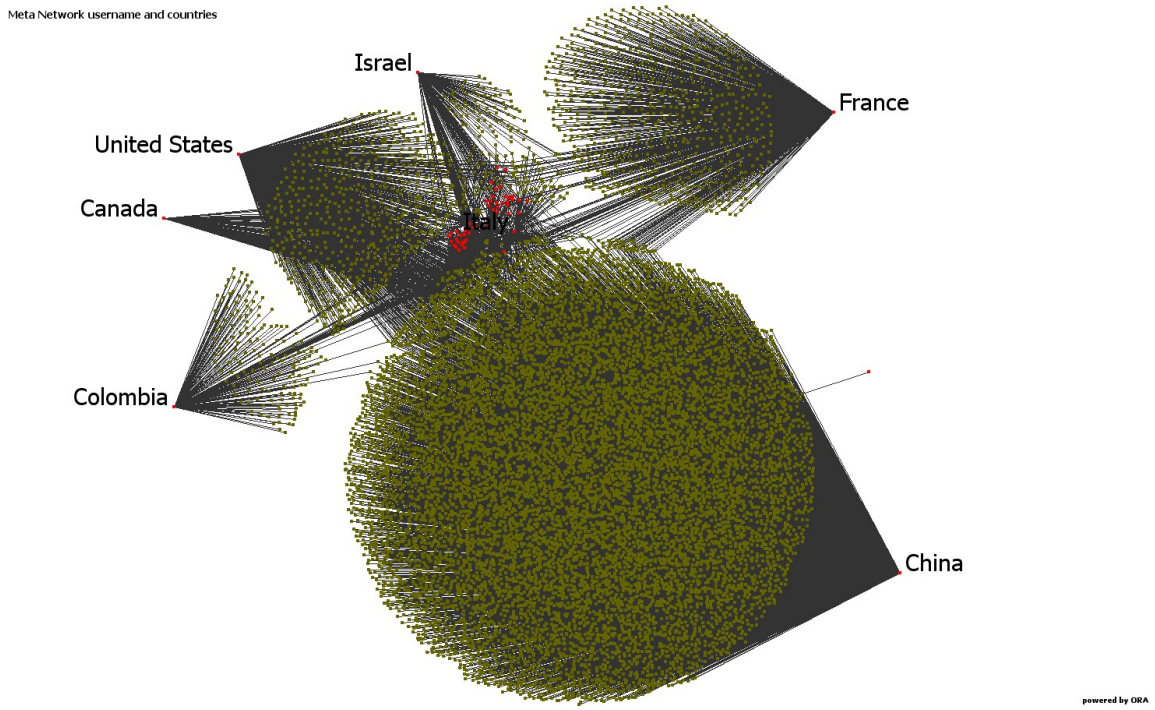
Graph 3. A network of countries and usernames.
(Orange dots are IPs addresses, and blue dots are usernames)



The two networks present the relations between IPs and UNs (Graph 3) and PWs (Graph 4). The networks present clear and defined subnetworks in the network, meaning that the network can be divided into groups that have a massive volume of inner links between them.

In Graph 3, it is very clear that (starting from the left corner) attacks coming from Colombia, Ukraine, the United States, Canada, France, China, and Israel use unique and specific words to guess usernames in their attacks. In Graph 4, it is evident that (starting from the left corner) attacks coming from Canada, the United States, Israel, France, China, and Colombia use unique and specific words to guess passwords in their attacks. In the middle of each graph are attacks coming from countries that share their UNs (in graph 3) and PWs (in graph 4) with many other countries.

Graph 4. A network of countries and passwords.
(Orange dots are IP addresses, and green dots are passwords)



To identify the existence of such groups and explore them, the study uses the Newman algorithm (Newman, 2004). This algorithm is a standard algorithm that aims to find distinct groups in a specific network that are included in the network and the members of each group. The algorithm is based on the number of links each node in the network has with all other nodes and based on the ratio of links each node has with its surrounding neighbors; the algorithms divide the network into specific groups with a high volume of reciprocal links (which are sometimes are called communities). Tables 1 and 2 present the results of applying this algorithm to the networks.

Table 1 presents the communities in the country-username network:

Table 1. Communities in the country-username network

Community	Countries	# Members
A	China	1
B	France	1
C	Colombia	1
D	USA and Canada	2
E	All the Rest	49

Table 2 presents the communities in the country-password network:

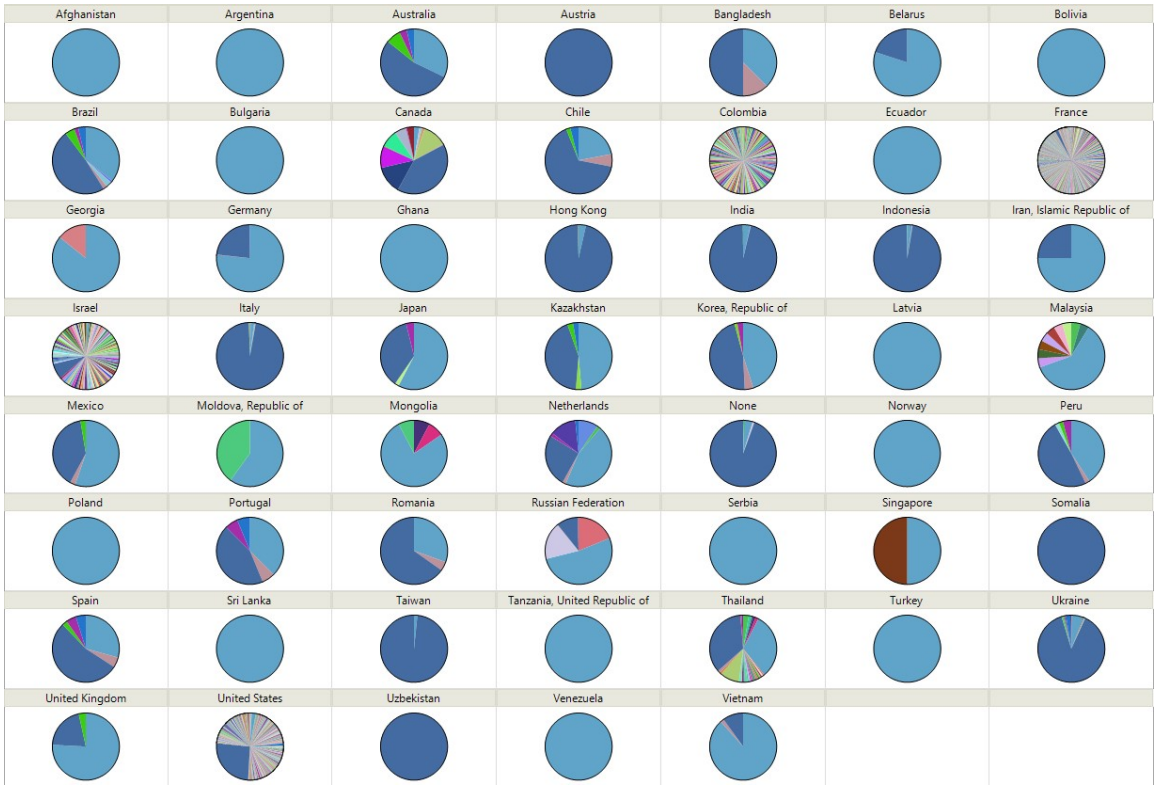
Table 2. Communities in the country-password network

Community	Countries	# Members
A	China	1
B	France	1
C	All the Rest	52

From Tables 1 and 3, it is clear that the mathematical analysis supports the visual analysis of the networks. The algorithm finds specific communities that are very similar to the defined communities presented in Graph 3 and Graph 4.

To better visualize the differences in the distribution of the usage of specific usernames, Figure 1 presents the distribution of various UNs initiated from different IP addresses.

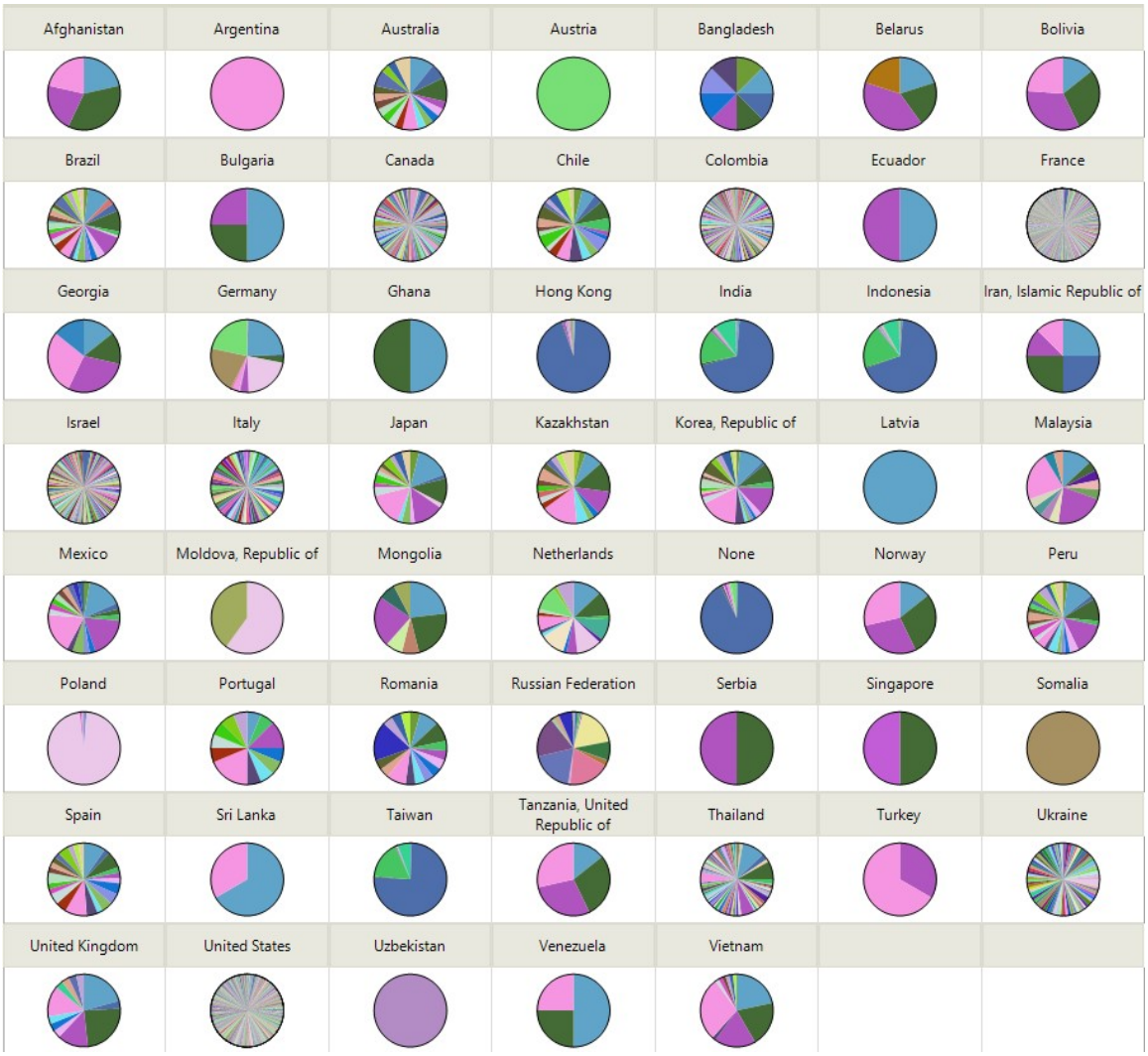
Figure 1. The distribution of multiple usernames initiated in IP addresses
(Light blue is= "Admin" and dark blue= "root," all other colors are less common user names).



Looking at Figure 1 from the left corner and going down each row, one can see that Canada, Columbia, France, Israel, and the US have the highest usage of a variety of UNs. Among all five countries mentioned above, the US is slightly different since almost a quarter of its attacks originated using a single username ("Admin").

Figure 2 presents the differences in the usage of different passwords.

Figure 2. The distribution of various passwords initiated in IP addresses
(Light pink is= "Admin", dark pink= "letmein", Light blue is= "1234",
dark blue= "123456", green is= "root", all other colors are less common passwords).



A similar trend can be found in Figure 2. Looking at Figure 2 from the left corner and going down each row, one can see that Canada, Columbia, France, Israel, and the US have the highest volume of attacks and use a variety of PWs. These five countries have the greatest variety in their use of PWs. While Ukraine and Italy also use a variety of PWs, their usage of UNs is quite restrained (both use mostly the "root" UN).

Next, Table 3 presents the countries with the highest number of unique pairs of UNs and PWs. The term "unique" in this context means that the combination of a specific UN and a specific PW was found only in BFAs originating from a single country; the study did not find the usage of this combination in the attacks of any other country.

For example, in BFAs coming from IPs in China, 11,367 pairs of UNs and PWs were used only in BFAs coming from China. None of these 11,367 pairs appeared in attacks that were initiated from other IPs (outside China).

Table 3. Top countries with unique pairs of usernames and passwords

Country	Number of unique pairs
China	11,367
France	747
United States	675
Colombia	175
Israel	139
Canada	115
Italy	24
Ukraine	24
Brazil	18
Thailand	17
Russian Federation	7

While not all the pairs in our dataset are unique and some appear in several BFAs coming from different IPs, the existence of so many unique pairs (see Table 3) in attacks coming from IPs located in specific countries strengthens our hypothesis that different words are used by various hackers operating from different countries (different IPs). However, a more interesting conclusion can be reached from the data. In case a rare UN and PW combination appears in several BFAs coming from different countries, either the same hacker is operating from different IPs, or several hackers use the same corpus of words. This finding can be a signal of relations between hacking communities, and it might signal the sharing of practices and tools.

C. Sharing Patterns

Next, and as a result of the findings in part B, the study explores the pairs of UNs and PWs that appear in different countries. Table 4 presents the results of the countries with the most shared word pairs.

Table 4. Top countries with shared pairs of UN and PW

Country	Number of shared pairs
United States	374
China	366
Canada	134
Ukraine	87
Thailand	72
Israel	67
Italy	66
France	65
Brazil	31
India	30

Looking at Table 3 and Table 4 makes it clear that the top countries in terms of unique word pairs and the top countries in terms of shared pairs of words are almost identical. However, the number of shared pairs is much smaller than the number of unique pairs, and in contrast to the clear leadership of China in the number of unique pairs (85%), most of the sharing is done by the US and China (20% each). Table 5 presents each country's sharing patterns.

Table 5. Top countries with unique and shared pairs in percentage

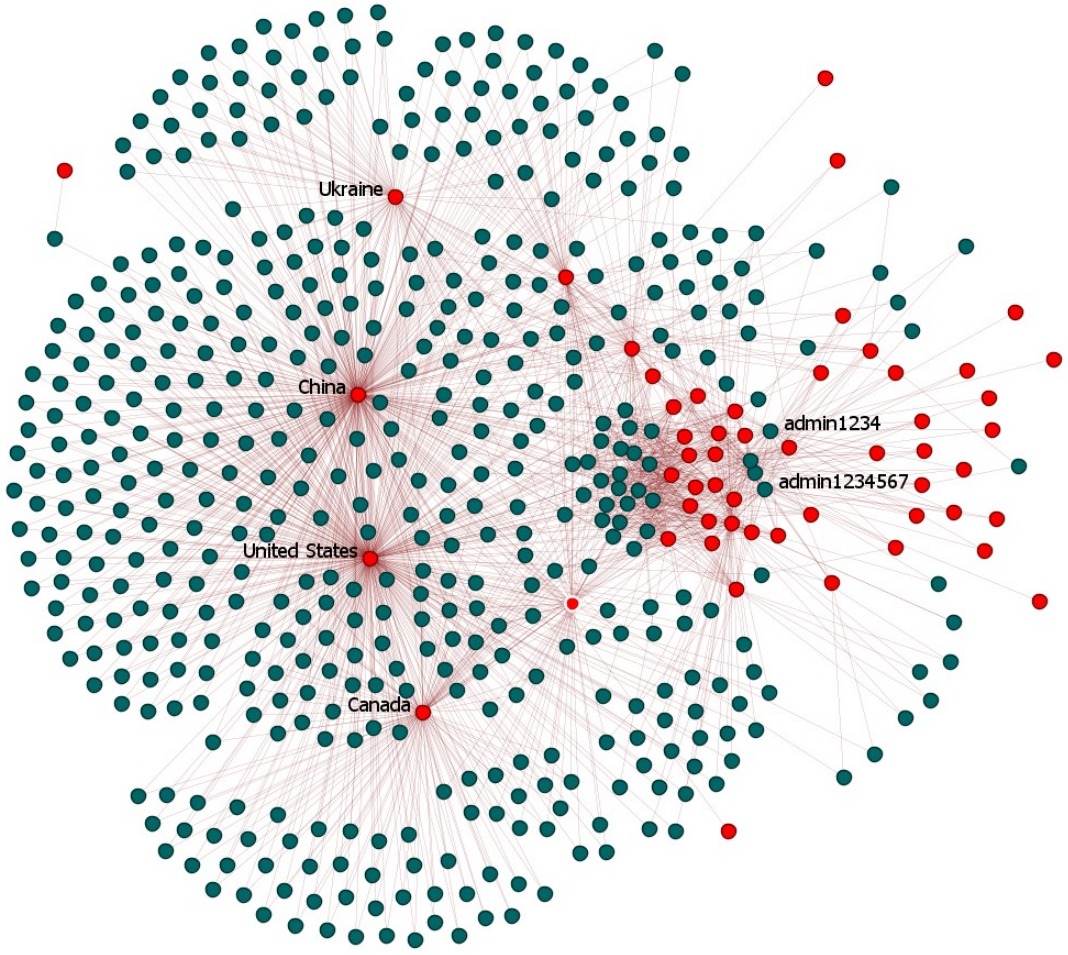
IPs Country	% shared from all shared pairs	% unique from all unique pairs	% shared from pairs of country	% unique from pairs of country
United States	21.2%	5.218%	36%	64%
China	20.7%	86.479%	3%	97%
Canada	7.6%	0.039%	96%	4%
Ukraine	4.9%	0.156%	81%	19%
Thailand	4.1%	0.078%	88%	12%
Israel	3.8%	0.903%	37%	63%
Italy	3.7%	0.171%	75%	25%
France	3.7%	5.476%	8%	92%
Brazil	1.8%	0.008%	97%	3%
India	1.7%	none	100%	0%
Indonesia	1.6%	0.016%	93%	7%
Colombia	1.5%	1.309%	14%	86%
Republic of Korea	1.5%	0.008%	96%	4%
Russian Federation	1.5%	0.047%	81%	19%
Peru	1.4%	0.008%	96%	4%

Table 5 presents three patterns of usage of unique UNs and PWS in BFAs. There are countries such as China, France, and Colombia, where almost all of the BFAs coming from their IPs use unique UN and PW pairs. In contrast, nearly all of the BFAs coming from Canada, Thailand, Ukraine, and Brazil use shared pairs of words. In addition, there are countries, such as the US and Israel, where the proportion of unique pairs and shared pairs is two to one.

Building and exploring the network of countries that share pairs of words can provide insights concerning the relations of these countries.

Graph 5 presents such a network.

Graph 5. Countries and pairs of usernames and passwords are shared in the BFAs.
(red dots are IPs, and Green dots are pairs of usernames and passwords)

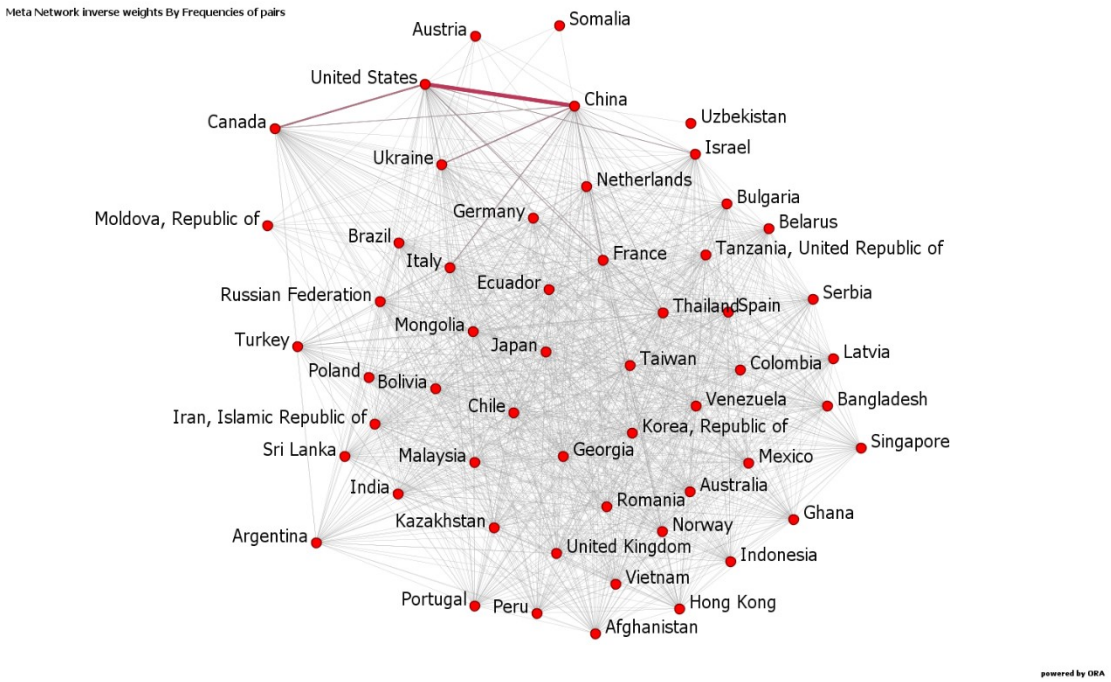


Graph 5 shows the pairs of words and the countries (IPs) that share them during the BFAs. The network is bipartite and has IPs and pairs as nodes. The central pair nodes (highest total centrality) that are shared are (not surprisingly) the words "Admin" as UN and "1234" or "1234567" as PWs. The central IP nodes (highest total centrality) are China, the US, Ukraine, and Canada.

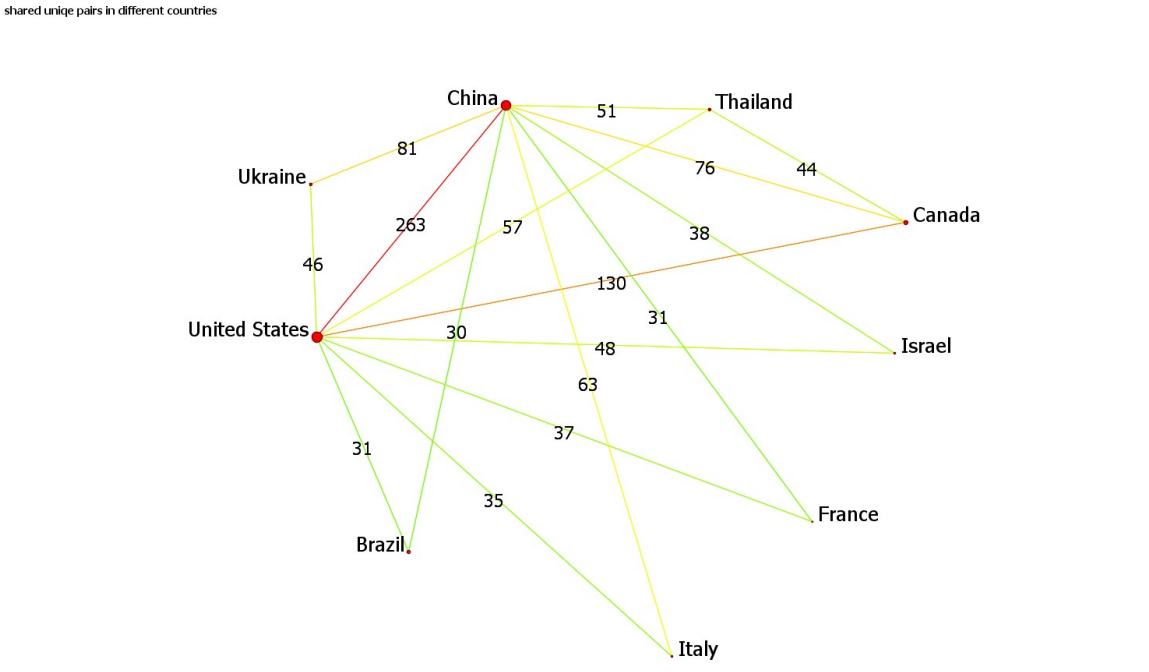
Based on this network, the study builds a network of only IPs. This network folds the pairs of words and turns them into links that connect IPs. The volume of the strength of a link between two specific IPs is the number of pairs that these two IPs have in common. Graph 6 presents this network.

Graph 6 presents all IPs that share UN and PW pairs with other IPs. One can see that almost every IP shares at least one UN and PW pair with another IP (in our dataset). Since it is a dense network, and the links are not easily seen, Graph 7 presents only IPs (countries) that share more than 30 unique pairs of usernames and passwords with other IPs (countries). (The threshold of 30 represents an average of one UN and PW pair shared per day).

Graph 6. Countries and pairs of usernames and passwords are shared in the BFAs
(red and thick lines are the links with high value,
meaning that more pairs of usernames and passwords are shared).



Graph 7. Links between IPs of countries that share more than
30 unique pairs of usernames or passwords.
(Red links have a higher number of shared pairs. The value is presented above the link).



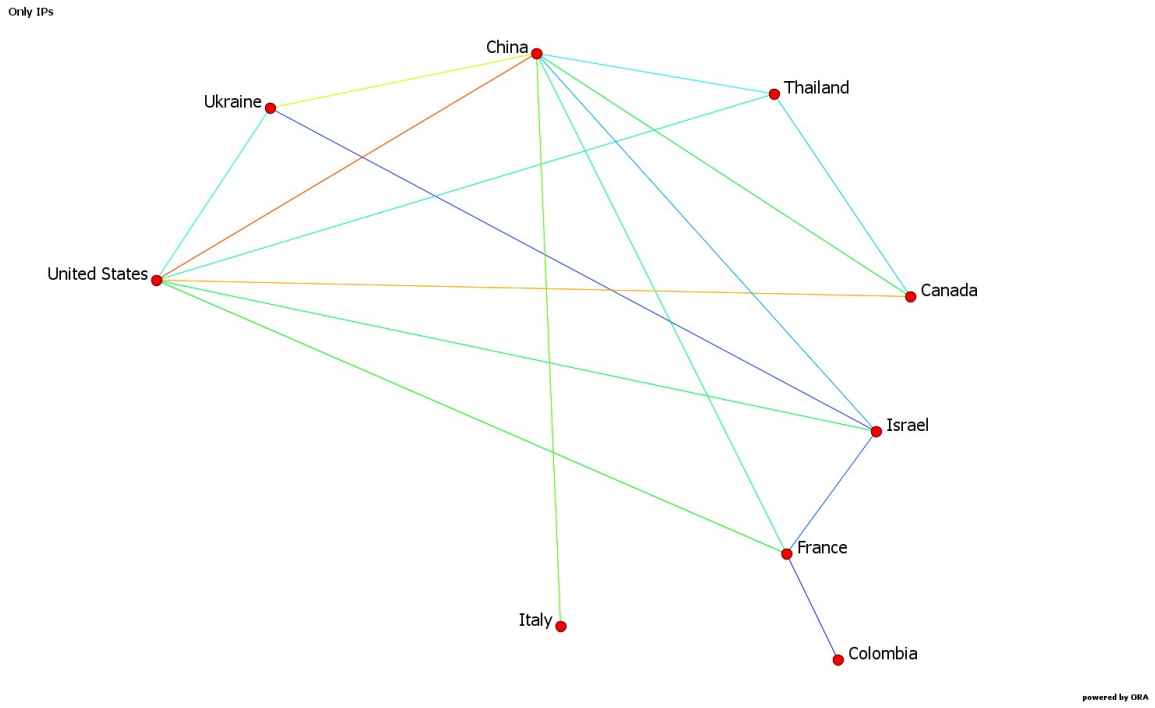
Various IPs share one thousand, seven hundred, sixty-six pairs of names and passwords. China and the US share almost 300 (17%), while the US and Canada share 130 (7.3%). China shares nearly the same number of pairs with Ukraine and Canada (81 and 76, respectively). On the other hand, Colombia, which initiates many BFAs, does not share more than 30 pairs with any other country, and it seems that Colombia has its own corpus of words.

To dive even deeper into the pattern of IP cooperation in BFAS, the study suggests that cooperation between IPs is better explained through the analysis of their mutual usage of infrequent pairs of UNs and PWs.

Thus, the utilization of rare combinations of username and password by two IPs is worth more (as a signal of their cooperation) than the usage of ubiquitous pairs (for example, UN= "Admin" and PW= "1234"). The weight of the links between IPs has to be calculated in correlation to the inverse of their frequencies to depict the importance of the usage of rare pairs. Graph 8 presents the network between the leading IPs where the strength of links is based on the frequency of the words that connect the IPs.

Graph 8. Links between IPs of countries where rare pairs obtain a higher weight

(Red links are made of rare words and are thus stronger links.
Blue links are made of common words and are therefore weaker links).



The researchers recalculated the strength of the tiers, including the higher weight of rare shared UN and PW pairs. A comparison between Graph 7 and Graph 8 reveals minor differences between them. IP identity is mostly the same (except for Brazil and Colombia), and the strength of links between IPs is almost the same. This means that BFAs coming from China and the US share many of the same words. BFAs coming from the US and Canada share many words, and the same is true for the BFAs coming from Ukraine and China.

Discussion and Conclusion

The goal of this study was to explore whether there are unique usernames and passwords that different hackers use and whether there is a spatial pattern of this usage.

Our data allow us to examine both failed and successful attempts of BFAs while exploring the data hackers used in their penetration attempts in an effort to contribute to the knowledge on hacker communities and information sharing.

The first findings of the analysis revealed that there are various monthly patterns of attacks and different daily patterns of attacks coming from different IPs. This finding implies the existence of distinct hacking processes, techniques, and probably tools that are being used in different countries. Detecting such patterns allows researchers and law enforcement agencies to better differentiate between hackers and groups of hackers or bots. This implies that different attackers work using not only various algorithms but also word corpora.

This understanding led to the second research question to explore that modus operandi and the data corpora.

The second finding reveals that for both usernames and passwords, there are various and unique corpora used by hackers (humans and robots) to execute brute force attacks. China, France, Colombia, the US, and Canada, use unique corpora of words in their BFAs.

An analysis of the cooperation of different IPs based on the usage of rare UNs and PWS did not uncover new patterns. BFAs using IPs from China, France, Colombia, the US, and Canada, do not collaborate much with BFAs coming from other countries (other IPs).

The findings suggest that closed groups of attack (initiated by one hacker from a specific group of hackers) share their corpora based on a spatial basis (IP location).

There are collaborations between spatial communities, but an analysis of the network suggests that unique corpora of words are still in use in each country.

The practical contribution of the study's findings is rooted in the suggested match between words used in BFAs and the origin of the hacker attacks.

Since a real technological difficulty exists in identifying the source of hacking attempts, the ability to connect the usage of certain word corpora to a suggested country of origin might help in tracking down hackers and hacking attempts.

BFAs might indeed come from a specific IP address; they might be initiated in other IPs. However, the significant result and the subnetwork found in the analysis support the hypothesis that there are hacker communities and information sharing based on geographical locations.

A third finding is that there is also a (limited) shared corpus of usernames and passwords that BFAs around the globe use. The shared corpus is an example of the hackerspaces (Moilanen, 2012) where hackers collaborate and share resources such as knowledge (Rechavi et al., 2015) and malware (Macdonald & Frank, 2017) around the globe. This finding does not contradict the second finding but rather enriches it. In a large network containing many hackers coming from different IPs, there are several forms of cooperation. Local communities who work on a spatial basis and share techniques and knowledge exist side by side with global communities that share several local hacking words, dictionaries, and knowledge.

This analysis suggests a new direction to disrupt hacking communities and networks. Such an analysis might enable law enforcement agencies to map the hacking network

better and differentiate hackers and hacker groups based on their use of words. Additionally, this type of analysis can also contribute to the identification of hackers' locations or, at the very least, the location from which they choose to hack.

Limitations and Future Research

Like all studies, this study is limited in its data. While hacking networks are dynamic, the findings may be specific to 2016—the year in which the study explored the network.

Moreover, the scope of the data is limited to an academic institution in China, and it might not reflect the process and range of hacking-word corpus usage. Educational sites might attract hackers whose interests are focused on research or educational materials. Their way of work and collaboration might be different from political hackers, financial hackers, or jack-of-all-trades hackers.

Despite its limitations, the study contributes to the knowledge on the crime script of the hacking community, especially in terms of the countries (IPs of BFAs) in which hackers share their work methods and expertise. Future studies should include online data coming from different sources, such as HPs located in varied geographical locations and various organizations or hacking forums.

Acknowledgments

This research was conducted with support from the Israeli Ministry of Science, Technology, and Space (Grant No. 3-10888) and by the Federmann Cyber Security Center in conjunction with the Israel national cyber directorate.

References

Barber, R. (2001). Hackers profiled: Who are they and what are their motivations? *Computer Fraud & Security*, 2001(2), 14-17.

Benjamin, V., & Chen, H. (2012, June). Securing cyberspace: Identifying key actors in hacker communities. In *2012 IEEE International Conference on Intelligence and Security Informatics* (pp. 24-29). IEEE.

Bonneau, J. (2012, May). The science of guessing: analyzing an anonymized corpus of 70 million passwords. *In 2012 IEEE Symposium on Security and Privacy* (pp. 538-552). IEEE.

Bouchard, M., & Amirault, J. (2013). Advances in research on illicit networks. *Global crime*, *14*(2-3), 119-122.

Bright, D., Greenhill, C., Britz, T., Ritter, A., & Morselli, C. (2017). Criminal network vulnerabilities and adaptations. *Global Crime*, *18*(4), 424-441.

Chandler, A. 1996. The changing definition and image of hackers in popular discourse. *International Journal of the Sociology of Law*, *24*(2), 229-251.

Chavez, N., & Bichler, G. (2019). Guarding against Cyber-Trespass and Theft: Routine Precautions from the Hacking Community. *International Journal of Cyber Criminology*, *13*(1).

Dave, K. T. (2013). Brute-force Attack "Seeking but Distressing". *International Journal of Innovations in Engineering and Technology*, *2*(3), 75-78.

Décary-Hétu, D., & Dupont, B. (2012). The social network of hackers. *Global Crime*, *13*(3), 160-175.

Dupont, B., Côté, A. M., Savine, C., & Décary-Hétu, D. (2016). The ecology of trust among hackers. *Global Crime*, *17*(2), 129-151.

Easton S. T., & Karaivanov A. K. (2009). Understanding optimal criminal networks, *Global Crime*, *10*(1-2), 41-65, doi: 10.1080/17440570902782444

Egelman, S., Sotirakopoulos, A., Muslukhov, I., Beznosov, K., & Herley, C. (2013, April). Does my password go up to eleven? the impact of password meters on password selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2379-2388). ACM.

Gagneja K., & Jaimes L.G. (2017) Computational Security and the Economics of Password Hacking. In: Doss R., Piramuthu S., Zhou W. (eds) Future Network Systems and Security. FNSS 2017. *Communications in Computer and Information Science*, vol 759. Springer, Cham

Holt, T. J. (2007). Subcultural evolution? Examining the influence of on- and off-line experiences on deviant subcultures. *Deviant Behavior*, *28*, 171–198.

Holt, T. J. (2013). Exploring the social organization and structure of stolen data markets. *Global Crime*, 14(2-3), 155-174.

Holt, T. J., & Kilger, M. (2012). Examining willingness to attack critical infrastructure online and offline. *Crime & Delinquency*, 58(5), 798-822. doi: 10.1177/0011128712452963.

Holt, T. J. (2005) "Hacks, Cracks, and Crime: An Examination of the Subculture and Social Organization of Computer Hackers". *Dissertations*. 616. https://irl.umsl.edu/dissertation/616

Holt, T. J., Strumsky, D., Smirnova, O., & Kilger, M. (2012). Examining the Social Networks of Malware Writers and Hackers. *International Journal of Cyber Criminology*, *6*(1).

Holt., T. J., & Heith., C. (2010) Transferring Subcultural Knowledge On-Line: Practices and Beliefs of Persistent Digital Pirates, *Deviant Behavior*, *31*(7), 625-654. doi: 10.1080/01639620903231548

Jordan, T., & Taylor, P. (1998). A Sociology of Hackers. *The Sociological Review*, *46*(4), 757–780. doi: 10.1111/1467-954X.00139

Jose, J., Tomy, T. T., Karunakaran, V., Krishna, A., Varkey, A., & Nisha, C. A. (2016, March). Securing passwords from dictionary attack with character-tree. In *2016 International Conference on Wireless Communications*, Signal Processing and Networking (WiSPNET) (pp. 2301-2307). IEEE.

Juels, A., & Rivest, R. L. (2013, November). Honeywords: Making password-cracking detectable. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (pp. 145-160). ACM.

Kelley, P. G., Komanduri, S., Mazurek, M. L., Shay, R., Vidas, T., Bauer, L., ... & Lopez, J. (2012, May). Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *2012 IEEE symposium on security and privacy* (pp. 523-537). IEEE.

Kleinknecht, S. W. (2003), Hacking Hackers: Ethnographic Insights into the Hacker Subculture-Definition, Ideology and Argot. Dissertation. Retrieved from http://hdl.handle.net/11375/10956

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, *52*(4), 667-694.

Leukfeldt, R., Veenstra, S., & Stol, W. (2013). High volume cybercrime and the organization of the police: The results of two empirical studies in the Netherlands. *International Journal of Cyber Criminology*, *7*(1), 1.

Lu, Y., Luo, X., Polgar, M., & Cao, Y. (2010). Social network analysis of a criminal hacker community. *Journal of Computer Information Systems*, *51*(2), 31-41.

Ma, J., Yang, W., Luo, M., & Li, N. (2014, May). A study of probabilistic password models. In *2014 IEEE Symposium on Security and Privacy* (pp. 689-704). IEEE.

Macdonald, M., & Frank, R. (2017). The network structure of malware development, deployment and distribution. *Global Crime*, *18*(1), 49-69

Madarie, R. (2017). Hackers' Motivations: Testing Schwartz's Theory of Motivational Types of Values in a Sample of Hackers. International *Journal of Cyber Criminology*, *11*(1).

Malm, A., Bichler, G., & Nash, R. (2011). Co-offending between criminal enterprise groups. *Global Crime*, *12*(2), 112-128.

Mazurek, M. L., Komanduri, S., Vidas, T., Bauer, L., Christin, N., Cranor, L. F., ... & Ur, B. (2013, November). Measuring password guessability for an entire university. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (pp. 173-186). ACM.

Moilanen, J. (2012) Emerging hackerspaces–peer-production generation. In IFIP International Conference on Open Source Systems (pp. 94-111). Springer, Berlin, Heidelberg.

Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*(6), 066133.

Olson, P. (2012). *We are anonymous: Inside the hacker world of LulzSec, anonymous, and the global cyber insurgency*. New York: Back Bay Books.

Pliam, J. O. (2000, December). On the incomparability of entropy and marginal guesswork in brute-force attacks. In: *International Conference on Cryptology in India* (pp. 67-79). Springer, Berlin, Heidelberg.

Raza, M., Iqbal, M., Sharif, M., & Haider, W. (2012). A survey of password attacks and comparative analysis on methods for secure authentication. *World Applied Sciences Journal*, *19*(4), 439-444.

Rechavi, A., Berenblum, T., & Maimon, D. (2018). The Secondary Global Market for Hacked Data. *International Journal of Cyber Criminology*, 12(2).

Rechavi, A., Berenblum, T., Maimon, D., & Sevilla, I. S. (2015, August). Hackers topology matter geography: Mapping the dynamics of repeated system trespassing events networks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* 2015 (pp. 795-804). ACM.

Sullivan, B. (2007). Preventing a brute force or dictionary attack: how to keep the brutes away from your loot. Pridobljeno (17.4. 2014) iz CODE Project: Retrieved from http://www.codeproject.com/Articles/17111/Preventing-a-Brute-Force-or-Dictionary-Attack-How.

Summers, T., (2015) How Hackers Think: A Mixed-Method Study of Mental Models and Cognitive Patterns of Hight-Tech Wizards. Dissertation. Retreived from https://etd.ohiolink.edu/!etd.send_file?accession=case1427809862&disposition=inline.

Summers, T., Lyytinen, K. J., Lingham, T., and Pierce, E. A. (2013) How Hackers Think: A Study of Cybersecurity Experts and Their Mental Models. Third Annual International Conference on Engaged Management Scholarship, Atlanta, Georgia. Paper 3.3

Thomas, D. 2002. *Hacker culture*. U of Minnesota Press.

Turgeman-Goldschmidt, O. (2005). Hackers' accounts hacking as a social entertainment. *Social Science Computer Review*, *23*(1), 8-23.

Von Lampe, K., & Ole Johansen, P. (2004). Organized Crime and Trust: On the conceptualization and empirical relevance of trust in the context of criminal networks. *Global Crime*, 6(2), 159-184.

Weir, M., Aggarwal, S., De Medeiros, B., & Glodek, B. (2009, May). Password cracking using probabilistic context-free grammars. *In 2009 30th IEEE Symposium on Security and Privacy* (pp. 391-405). IEEE.

Zhang, X., Tsang, A., Yue, W.T., and Chau, M. (2015) The classification of hackers by knowledge exchange behaviors. *Information Systems Frontiers*, 17(6), pp.1239-1251.

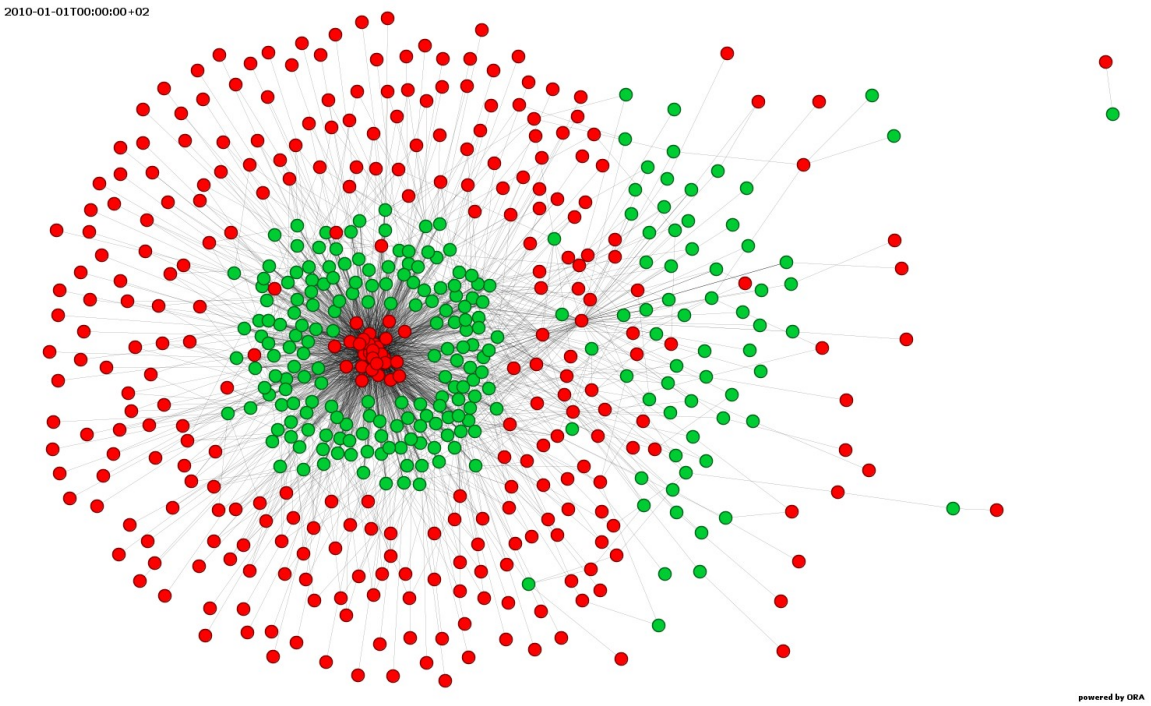
Zhang-Kennedy, L., Chiasson, S., & Biddle, R. (2013, September). Password advice shouldn't be boring: Visualizing password guessing attacks. In 2013 APWG eCrime Researchers Summit (pp. 1-11). IEEE.

Appendix A - The hackers' IPs and the honeypot network

Graph A1 presents the relations between the hackers' IPs and the HPs.

Graph A1- The relationship between IPs and HP

(green dots are HPs, and red dots are IPs initiating BFAs).



Graph A1 presents the sum of the attacks during the 35 days. Many IPs attack most of the HPs, and only a few IPs attack specific HPs. The red dots in the center of the network are IPs that attack almost all HPs. These are the most active IPs, and in the sense of quantity, they are the most harmful IPs.

The study does not analyze this network, and it is presented here to provide a rough understanding of the relations between IPs and HPs.

Appendix B - The distribution of words per country

Since each BFA includes a username and password, this study explores the usage of usernames and passwords in all BFAs. There were 1,552 usernames and 12,975 passwords. Since China is an outlier, Table B1 presents the top 10 combinations of usernames and passwords in China.

Table B1. Leading username and password combinations in attacks coming from IPs in China

Origin	Username	Password	Frequency
China	Root	123456	7545
China	Root	password	2033
China	Root	1qaZ2wsX	1955
China	Root	1Q2w3e4r	1843
China	Root	admin	1706
China	Root	root	1507
China	Root	root123	1487
China	Root	123456789	1382
China	Root	abc123	1314
China	Root	P@ssw0rd	1295

Table B2 presents the top 10 username and password combinations in other countries (excluding China).

Table B2. Leading username and password combinations in attacks coming from IPs outside China

Origin	Username	Password	Frequencies
India	Root	123456	1694
Indonesia	Root	123456	1413
Taiwan	Root	123456	1168
Hong Kong	Root	123456	396
Indonesia	Root	123456	395
India	Root	123456789	377
Ukraine	Root	admin	273
Taiwan	Root	123456789	260
India	Root	qwerty	228
Ukraine	Root	password	216