2015 10th Asia Joint Conference on Information Security

The Cryptanalysis of WPA & WPA2 in the Rule-Based Brute Force Attack, An Advanced and Efficient Method

Chia-Mei Chen
Department of Information Management
National Dr. Sun Yat-sen University
Kaohsiung, Taiwan
cmchen@mis.nsysu.edu.tw

Tien-Ho Chang
Department of Information Management
National Dr. Sun Yat-sen University
Kaohsiung, Taiwan
d024020001@student.nsysu.edu.tw

***Abstract*—The development of kinds of mobile device is a nonlinear but in a tremendous hopping way. The security of wireless LAN is far more important, and its mainly present protection is the WPA & WPA2 protocol which is a complex tough algorithm. This exploratory study shows that there is a security gap by the social human factors which are the weak passwords. Traditionally, brute force password attack is using the dictionary files that is aimless and extremely labor work. Now, we proposed 10 rule-based methods which are globally inclusive and culturally exclusive and prove the insecurity of WPA & WPA2 by 100 empirical and valuable real wireless encrypted packets of WPA & WPA2. The evidence shows that there is a 68 % of cracking rate and then do the passwords patterns analysis as well.**

Keywords—cryptanalysis; WPA & WPA2; rule-based; wireless security; brute force attack; dictionary attack

I. INTRODUCTION

The development and purchase of mobile device using the wireless LAN communication is more and more vigorous. According to the IDC (International Data Corporation), vendors shipped a total of 327.6 million units during the third quarter of 2014 (3Q14), resulting in 25.2% growth when compared to the 261.7 million units shipped in 3Q13 and 8.7% sequential growth above the 301.3 million units shipped in 2Q14 [1]. The worldwide tablet grew 11.0% year over year in the second quarter of 2014 (2Q14) with shipments reaching 49.3 million units [2]. The combined consumer and enterprise worldwide wireless local area network (WLAN) market segments grew 9.2% year over year in the second quarter of 2014 (2Q14) [3]. By the data of Wigle.net [4], there are 122.37 million unique Wi-Fi networks on 1st Feb. 2014, and 173.54 million unique Wi-Fi networks on 3rd Feb. 2015 which is the growth of 41.8% in one year. From those statistics above, we can see that the growing usage and shipment of mobile device is getting prosperous and the trend is irreversible, and its security issues followed would be easily ignored for convenience.

Along with the advancing times and the technological development, WLAN technology has been widely applied in our daily work [5]. Since the standard of 802.11 for wireless LAN addressed in 1997, it has been the stages of a, b, g, n, and ac that makes the faster speed, longer distance and more stable quality with the MIMO (Multi-input Multi-output) antenna, and there is also the 5G wireless LAN which enhances the overall more excellent transmission ability than the 2.4G access point. With that excellence of transmission capacity, there are all kinds of various wireless LAN activities with the tremendous download volume of 50 billion for Google play or Apple Market until 2Q13. In short, the hardware of and software within the applications in wireless made the secure issues inevitable and no way to dodge, and the most critical part is the encryption of wireless packets which contain kinds of useful information, such as passwords, causing the major secure problem. Especially, the most important of all is the passwords of the mobile devices which means the full control of the devices, and the mainstream encryption of wireless LAN is the protocol of WPA & WPA2 which is the target of this research. The Complication of WPA &WPA2 is well designed which is hashed 4096 times per encryption [6], and its algorithm calculation is irreversible, that is, the only way to crack it is comparing the PMKs (the Pairwise Master Key). Basically, the encryption of WPA & WPA2 is mainly the calculation of SSID and password of wireless Access Point, and finally got the value of PMK at which is the aim of the malicious attackers.

Traditionally, the methods of cracking WPA & WPA2 passwords is mainly the implementation of brute force dictionary attack [7] because the way of handling the algorithm is impractical, and with that type of attack is also named the dictionary attack which contains numerous password candidates inside the files, that's extremely inefficient. But it is still the common and practical way to crack those passwords by brute force comparing method. The relationships between the cryptanalysis of WPA & WPA2 and brute force attack is so close, and we need to do the dictionary generation for the password candidates and the subsequent file size is terrible statistically with the 4096 encrypted processes for single password candidate. As for the protocol of encryption, from our wardriving data, we can clearly tell that the mainstream of encryption in wireless LAN from the street is WPA & WPA2. And more, the capturing of wireless encrypted packets get its

978-1-4799-1989-5/15 $31.00 © 2015 IEEE
DOI 10.1109/AsiaJCIS.2015.14

own difficulty from the streets which is perhaps not mentioned in the previous literature.

We regard this research as an exploratory research which the way the cryptanalysis of WPA & WPA2 by using rule-based method without generating any dictionary files or using any dictionary files. The way we do the implementation of decryption analysis is “clean” and “parsimony” because it is always time-consuming, energy-exhausted, and endless huge volume of hard disk capacity underlying the whole process of traditional dictionary attack. Owing to the humongous combination of 95^{63} max in the cryptanalysis of WPA & WPA2, we primary focus mainly on the number digit and partly alphabets for culturally exclusively and globally inclusively and it perhaps firstly testify the real security of WPA & WPA2 on the streets which the places we collect the real four way handshake wireless encrypted packets. With the aid of the GPU-based computing in cryptanalysis, our rule-based method reached a 68% cracking rates without generating any dictionary files showing its parsimony, and didn’t compare to the traditional brute force dictionary which is computing passwords one by one with the 4096 times per hash value. Furthermore, it’s extremely aimless in selecting the content of a dictionary file or generating the file which would yield to a humongous disk volume.

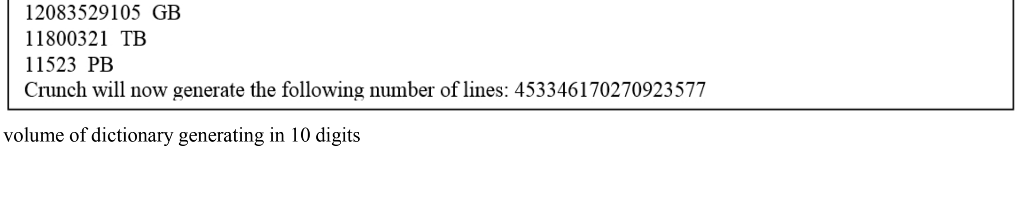
II. The Cryptanalysis of WPA &WPA2 and Dictionary Files

A. Brute force and Dictionary Files in Cracking WPA & WPA2

A dictionary attack is a password cracking method in which every single word from a word list is tried [8]. Before getting insight into the cracking passwords of the WPA & WPA2, we have to get the basic picture of how the encryption dealt with between the AP and its clients and the procedure of capturing the encrypted packets for testing which is one of the key elements in WPA & WPA2 cryptanalysis. We take the real samples from the street, in contrast to the lab experiment setting, which might cause certain degrees of difficulties in sampling. Basically, the mechanism between the AP and its clients in encryption and authentication is called four way handshake which is the peculiar phrase we could get the encrypted packets for cryptanalysis. Once, we got the target packets, what we have to do is the comparing the PMK values of that target one, and cracked it if it meets the value. Traditionally, researcher always do the comparing by the brute force method, except the algorithm which is the most tough way in handling the cryptanalysis of WPA & WPA2, and consequently comes the issue of the dictionary files. First, the brute force method in passwords analysis is basically doing the labor work comparing the certain values one by one, and what the particular special for the cryptanalysis of WPA & WPA2 is its 8 digits password candidates minimum and the extremely complicated algorithm of 4096 times per password encryption. Considering another fatal factor is the dictionary files needed for the brute force method, the volume of hard disk would be a huge monster because the password candidates of WPA & WPA2 is from 8 to 93 with a total of 95^{8} ~ 95^{63} combinations. For reason above, the brute force method in WPA & WPA2 is terrific time-consuming, hard disk volume-wasting, and impractical and that’s what the protocol of WPA & WPA2 addressed it a highly secure way of encryption. Though various difficulties faced in cryptanalysis, the brute force attack is still the appropriate way for that, of course, there are specific procedures proposed for the improvement and efficiency of cryptanalysis.

B. The Difficulties of WPA & WPA2 Cryptanalysis in Password Combinations

Differs from the usual research, almost the data is from the controlled experiment laboratory which might not reflect the real security situations. The first step in cryptanalysis is the sampling which is from the street with certain degree of difficulty in capturing the encrypted packets. The main concern in capturing the empirical sample packets is that there must be active connections between the AP and its clients which is the most difficult part. After capturing the real sample from the street, it runs into the major obstacle of WPA &WPA2 cryptanalysis, that is, the computational ability in decrypting the irreversible PMK values, and the usual i7 PC with 8 cores (3.4Ghz/core) got the total computing ability of PMKs/s is only around 5000 which would take 333.3 mins in the minimum 8 digits only-100 million. This minimum combinations of 100 million have already took that much time, and the rest more complex passwords combinations would yield to the unimaginable time-consuming for the max of 95^{63}. And the inefficient computational ability is also impractical in the generation of the dictionary files because the encryption of WPA & WPA2 in its complication is 4096 times per calculation. It is inevitable that the brute force attack must be something to do with the dictionary file of hard disk volume-taking which is around GB to PB with 10 digits password attack with all 95 password candidates as shown in Fig. 2. The most current research is the Lab environment which is in a slow speed of computing and limited range of passwords in 8 digit numbers and for its easiness of cryptanalysis, the password in Lab would usually set like 12345678 that we would never know the real situations of WPA & WPA2 security. Usually, the dictionary file of lab is relatively small around few to hundreds of Megabytes only, and the password candidates are about few thousands to hundred millions. As the limits mentioned above, those statistics implemented in current lab research studies is



```
root@bt: ~/ crunch -3.4#    . / crunch  10  10  -f  charset.lst  mixalpha – numeric – all – space – o  /root/test.txt
Crunch will now generate the following amount of data: 1297459058238249O315 bytes
12373533804304  MB
12083529105  GB
11800321  TB
11523  PB
Crunch will now generate the following number of lines: 45334617027O923577
```

Fig. 1. The files volume of dictionary generating in 10 digits

extremely relative small comparing the incredible passwords combinations of 95^{63} max.

C. The Social Human Factor of Password Candidates

Conventionally, we presumed that people do use the passwords for their kinds of mobile devices by its own convenience and considered that these are the strong types of passwords, such as [birth year + Date], [Cell Phone Num.], and [birthdate] + [Name] et al.. This social human factor in the setting of WPA & WPA2 passwords is a broad way for the malicious attackers which would make the invasion propagation easier. In this study, we would like to testify this point that the wireless LAN users would take the easy passwords, and grab the encrypted packets from the street for the empirical evidence which not mentioned in the previous studies. Furthermore, we would like to propose an advanced, efficient, and rule-based method in cryptanalysis of WPA & WPA2 which is contrast to the traditional or lab experiment settings. From that efficient method in cracking passwords, we expect to reach these points which save certain amount of time comparing with the previous studies and prove that weak passwords the most users took.

In sum, our primary focus is on the social human factor of the cryptanalysis of WPA & WPA2 which is the weakest part of this highly secure algorithm, and we would like to prove our stand. We plan to propose the method of eliminating the impossible passwords combinations to the extent which is culturally exclusive and globally inclusive. Here, we implement several exclusive and inclusively rules based on Taiwan folk customs with some English usages about birth year + date et al.. With the appropriate exclusion of certain amount of nonsense password combinations, we could increase the efficiency of the cryptanalysis because the password candidates of WPA &WAP2 are $95^{8} \sim 95^{63}$, and no one would like to set the passwords like this “sdG&X1dJVr !8deqoA”. So that’s why we focus on the “lazy, weak, and convenient” passwords for cryptanalysis. Reference [9] showed that the efficiency improved by 14 times speedup with the exclusion of the nonsense phonotactics of Slovak using the statistical way by the John the ripper, but it is still with the annoying dictionary files with 14,572,531,443 lines of password candidates tested which would take certain vast amount of hard disk volume with the limited in the dictionary of a given language.

III. The Research Design

Traditionally, dictionary attacks are lucky picks by endless time [8]. Owing to the limits of speed in computing, the inefficiency of brute force dictionary attack [10] by calculating the nonsense password combinations with vast wasting volume of hard disk in cryptanalysis of WPA & WPA2. Here, we would like to propose a method containing 10 rules with the following advantages: 1st doing the cryptanalysis by GPUs without any dictionary files, 2nd culturally exclusive and globally inclusive in password candidates for efficiency under the complementary aid with the advanced cryptanalysis speed of 270,000 PMKs/s. This would make the better cracking rate expected, and show parsimony of the complicated cryptanalysis of WPA & WPA2, especially those with the tremendous large dictionary files. In our method, certainly, there is no a single byte produced or left in the whole process of cryptanalysis. We plan to implement the empirically collected wireless encrypted packets from streets to testify our proposed method. The wireless encrypted packet samples collected from the streets are absent from the present literature which could truly reveal the insecurity of WPA & WAP2 encryption by the human social factors. In order to create a statistical table which describes a mathematical model of a given language, it is necessary to build a dictionary which contains as many words of that language as possible [9]. In contrast, we propose an way of specific target-oriented password candidates choosing without using and generating any dictionary files, and its research design is as Fig. 2.

The software we implemented is the oclhashcat-plus [11] which is a free software, noncommercial and its algorithm is based on the Markov Chains which is a tool in various kinds of password cracking including combinator attack, dictionary attack, hybrid attack, mask attack, and rule-based attack. As we define the rule, then, the whole decryption computing would be put in the GPUs with mass parallel calculation (the GCN-Graphic Core Next architecture) [12] without any dictionary files generating. The rules we proposed are as the following:

A. Built-in Charset

From the following, we can find that there are 95 candidates for passwords and the length of WPA & WPA2 security protocol is 8 to 63 digits, so the total password combinations is $95^{8} \sim 95^{63}$. Each of these passwords would be hashed into encrypted values and each single password encrypted is by 4096 times of calculation which is the most difficult part of cryptanalysis. That’s why we focus on the certain range of password candidates in specific rules. And we plan to propose 5 categories: 8 digit numbers, local phone numbers, cell phone numbers, birthday format, and alphabets + numbers in 10 rules. Below, we list the basic built in charset for cryptanalysis, and propose the hypothesis that this is the

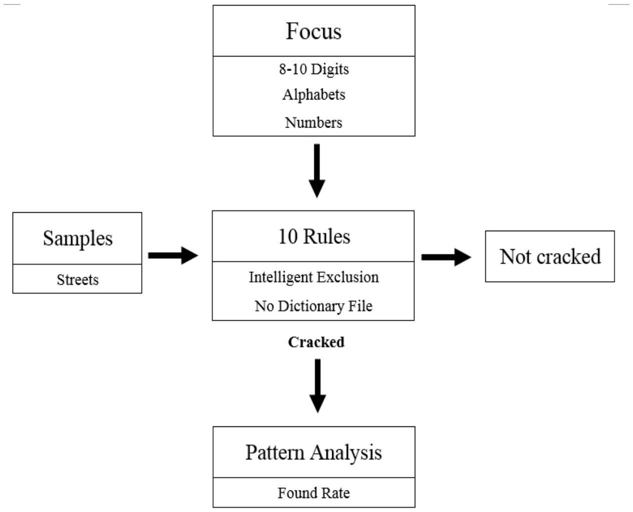


Fig. 2. The reseach design

greatly implemented zone in the password choices.

- ?l = abcdefghijklmnopqrstuvwxyz (lower case letters)
- ?u = ABCDEFGHIJKLMNOPQRSTUVWXYZ (upper case letters)
- ?d = 0123456789 (numbers)
- ?s = <space> !"#$%&'()*+,-./:;<=>?@[\]^_`{|}~ (symbols)

B. Mask Attack-Alphabets and Digits combination

We use the mask which designate the range to limit the variations with the certain target passwords in cryptanalysis without aimless cracking, especially the impossible composition of passwords by social human factors. Here, we focus on primary the 8 digits in mainly numbers and partly alphabets. With our outstanding parallel cracking speed-270,000 PMKs/s, the most easy one: ?d ?d ?d ?d ?d ?d ?d ?d, 8 digits number would take only 6.17 minutes to run the whole 100 million combinations.

- Cell phone
  - 09?d?d?d?d?d?d?d?d
- Numbers and alphabets
  - ?d ?d ?d ?d ?d ?d ?d ?d (ex. 12345678)
  - ?l ?d ?d ?d ?d ?d ?d ?d (ex., a1234567)
  - ?l ?l ?d ?d ?d ?d ?d ?d (ex., ab123456)
  - ?l ?l ?l ?d ?d ?d ?d ?d (ex., abc12345)
  - ?l ?l ?l ?l ?d ?d ?d ?d (ex., abcd1234)
  - ?u ?l ?d ?d ?d ?d ?d ?d (ex., Ab123456)

C. Rule-Based (Hybrid) Attack

Here, we focus on the numbers in 8 digits with the focus of local, cell phone number and birthday format, and the following is the composition of the rules which have four major parts, and each part of password candidates could be defined by users.

- -1 --custom-charset1= (?l, ?u, ?d, ?s)
- -2 --custom-charset2= (?l, ?u, ?d, ?s)
- -3 --custom-charset3= (?l, ?u, ?d, ?s)
- -4 --custom-charset4= (?l, ?u, ?d, ?s)

- Birthday format:

As for the birthday of 1900s, we set the rule as below which consists of the date from 19000101 to 19991231, and the beginning of 19 is fixed following the 2 digits: ?d?d, the months rule is: ?1?d (1: one), and the days are set as ?2?d. Thus, we can greatly reduce the time in cryptanalysis.

- -1 --custom-charset1=01
- -2 --custom-charset1=0123
- 19?d?d?1?d?2?d (1: one)

About the 2000s, we set the rule from 20000101 to 20151231 as depicted below, and the beginning of 20 is fixed following the sub rules 1 and 2, the months rule is: ?1?d, and the days rule is ?1?3.

- -1 --custom-charset1=01
- -2 --custom-charset2=012345
- -3 --custom-charset3=0123
- 20?1?2?1?d?1?3 (1: one)

- Local phone

It is extremely fast for cryptanalysis in local phone numbers because the certain areas are with the same beginning numbers, such as the rule set for author's local area-the beginning is 0755, and the rest digits are only 5, the rule could be as followed:

- -1 --custom-charset1=58
- -2 --custom-charset2=123578
- 075?1?2?d?d?d?d (1: one)

Owing to the factor of neighborhood, the victims who take the local phone number as the password of WPA & WP2 is especial extremely vulnerable because the combinations is greatly reduced by such rules and might be cracked in few seconds with our speeding GPUs (the GCN-Graphic Core Next architecture). Of course, the one could implement specific rules for their own needs with the targets, and our rules are based on the Taiwan folk conditions mostly.

D. Empirical Encrypted Wireless Packets Testifying

On the basis of the rules mentioned in previous sections, we collected 100 real wireless encrypted packets of four way handshake-WPA & WPA2 from the streets to verify our hypothesis of its insecurity of wireless networks. The data are valuable and difficult to collect and perhaps not mentioned in literature before. Combining the 10 rules proposed and the speeding GPUs (the GCN-Graphic Core Next architecture), we would take an insight into the secret of the complex highly secure and mainly secure protocol of WPA & WPA2, and we presume that the users would take the easy and convenient passwords for social human factors.

IV. Discussion and Conclusion

As an exploratory research, we just focus on the certain range of password candidates, especially those complex algorithm of WPA &WAP2 to testify its real street secure conditions, here, focus mainly on digit numbers and partly alphabets. To tell the specific principles, we could save the time and improve the efficiency to some extent by exclusion of the impossible passwords with social human factor, and more, don't have to handle the dictionary files or generating them with endless time and aimless labor work.

As for the principles of globally inclusive and culturally exclusive, this research proposes 10 rules which differentiate from the traditional dictionary brute force attack, on the contrary, we narrow the range of the target in a purposeful way of picking up the password candidates. Traditionally, the dictionary password attack is highly based on the dictionary files which is extremely not suitable for WPA & WPA2 protocol with its complex algorithm, and more, the choices and contents of certain dictionary is highly aimless or with huge amount of hard disk volume for the 95^{8} ~ 95^{63} of WPA & WPA2 password candidates.

From our data, we could see that there is a 68 % of cracking rate which is perhaps not proving in literature before, and do the password patterns analysis to see what the real situations of wireless networks security is. From Table I, we can tell the password patterns by our five categories in 10 rules and most of the users took the 1st and 3rd categories with the 70.5 % of the cracked passwords. It is that the minimum password input is 8 digits for WPA & WPA2 and the easiest way to fill is the simple human relative numbers for the convenient of remembering them. As for the 32.3 % of cell phone numbers, it is probably something to do with mobile device and most of the handy choice is its cell phone numbers which would be not easily forgotten at all. The rest fewer rate of local phone numbers, birthday format, and alphabets and numbers are relatively not so convenient as 8 digit numbers and the cell phone numbers. The lower rate of local phone number is perhaps not handy as the prevalent as cell phone numbers and most of the local phone number is only 7 digits without 2 digits local area number. And the birthday format, it might be that the anno domini (19xx) is different from Taiwan years with the 11 years gap. The final one is the alphabets and numbers is mostly not handy with 2 types of forms in the password inputs (alphabet and number).

From these empirical data, we can see that it fits our hypothesis that the users would take the most convenient way for password inputs by the social human factors. Though the encryption of WPA &WPA is tough, it still be vulnerable under the conditions of weak password and simplified password combinations linguistically and culturally. Our proposed method could be generate to all over the world by specific modification for one's own needs without producing any annoying dictionary files and target-oriented range of password candidates in the cryptanalysis of WPA & WPA2. By this research, we would like to remind the wireless LAN users of their insecure aspects of password choices, and provide the true evidence for researchers to see how vulnerable our real world of wireless security is. For the ethical regulations, this study obeys the laws of Taiwan which allows specific sniffing the wireless radio wave with the academic usage. We keep the data secret ,no logging in, no data theft, and no denial of service.

In the future, we plan to implement more targeted and rule-based password candidates picking in cryptanalysis, and extend our 8 digits to the more ones. Now, we are sorting out the phonotactics of certain the Chinese pin-yin (vowels and consonants sequences) and the common English spelling in Taiwan, preparing to combine these phonotactics with the digit numbers together for more efficient and parsimony of cryptanalysis of WPA & WPA2 for its complex algorithm.

TABLE I. CRACKED PASSWORDS PATTERNS

Patterns Analysis of Passwords		
Types of passwords	***Frequency***	***Rate***
8 digits numbers	26	38.23%
Local phone numbers	10	14.7%
Cell phone numbers	22	32.35%
Birthday format	6	8.82%
Alphabets and numbers	4	5.88%
Total	68	100%

REFERENCES

[1] "Mobile Device Growth 2Q14." [Online]. Available: http://www.idc.com/getdoc.jsp?containerId=prUS25224914.

[2] "Tablet Shippment Growth 2Q14." [Online]. Available: http://www.idc.com/getdoc.jsp?containerId=prUS25008314.

[3] "WLAN Market Growth 2Q14." [Online]. Available: http://www.idc.com/getdoc.jsp?containerId=prUS25077714.

[4] Wigle.net, "General Stats," 2015. [Online]. Available: https://wigle.net/gps/gps/main/stats/.

[5] L. Zhang, J. Yu, R. Zong, J. Chang, and J. Xue, "Prevention research of cracking WPA-PSK key based on GPU," in Proceedings of the 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), 2012, pp. 1965–1969.

[6] I. P. Mavridis, a. I. E. Androulakis, a. B. Halkias, and P. Mylonas, "Real-life paradigms of wireless network security attacks," in Proceedings - 2011 Panhellenic Conference on Informatics, PCI 2011, 2011, pp. 112–116.

[7] K. Tran, "GPU - accelerated WPA PSK cracking solutions," Minnesota State University, 2010.

[8] V. Bhatia, D. Gupta, and S. H.P., "Analysis of Dictionary Attack on Wireless Lan for Different Nodes," J. Inf. Syst. Commun., vol. 3, no. 1, pp. 167–169, 2012.

[9] J. Krekan, M. Pleva, and L. Dobos, "Statistical models based password candidates generation for specified language used in wireless LAN security audit," in Proceedings of the 20th International Conference on Systems, Signals and Image Processing (IWSSIP), 2013, pp. 95–98.

[10] Z. Jin, Y. Liu, and Y. Wang, "Survey on Security Scheme and Attacking Methods of WPA/WPA2," in Proceedings of the 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)., 2010, pp. 1–4.

[11] "Cyrptanalysis Software." [Online]. Available: http://hashcat.net/oclhashcat/.

[12] N. Nishikawa, K. Iwai, H. Tanaka, and T. Kurokawa, "Throughput and Power Efficiency Evaluations of Block Ciphers on Kepler and GCN GPUs," in Proceedings of the First International Symposium on Computing and Networking, 2013, pp. 366–372.