A Survey of Honeypot Research: Trends and Opportunities

Ronald M. Campbell
School of Computing
University of South Africa
Johannesburg, South Africa

Keshnee Padayachee
Institute for Science and Technology Education
University of South Africa
Pretoria, South Africa

Themba Masombuka
School of Computing
University of South Africa
Johannesburg, South Africa

***Abstract*—The number of devices connected to computer networks is increasing daily, and so is the number of network-based attacks. A honeypot is a system trap that is set to act against unauthorised use of information systems. The objective of this study was to survey the emergent trends in extant honeypot research with the aims of contributing to the knowledge gaps in the honeypot environment. The relevant literature was identified from a myriad of sources, such as books, journal articles, reports, et cetera. The findings suggest that honeypots are attracting the interest of researchers as a valuable security technique that can be implemented to mitigate network attacks and provides an opportunity to learn more about the nature of these attacks. Consequently a honeypot can be used as a research tool to gather data about network attacks.**

Keywords-Network security; Honeypots

I. INTRODUCTION

The number of devices connected to computer networks is growing rapidly. This has led to an increase in the number of network-based attacks. Kaspersky Labs [1] estimates these attacks at approximately 580 million. With such a vast number, implementing defence mechanisms to identify these attacks is important. One such defence mechanism is the use of honeypots. A honeypot in this study is viewed as a defence strategy for monitoring suspicious behaviour and recording proof of wrongdoing. A honeypot is a fake system that is put in place in order to fool intruders into believing that it is a real system. Unauthorised attempts to access the information are recorded by a honeypot with the aim of learning the techniques used by the intruder to gain access. Today, almost everyone is connected to a computer network, whether it is via computers or mobile devices like smart phones. People connect to these networks for various reasons, including conducting business transactions or being in touch with their friends and families using social networking sites. Many companies have realised this opportunity and are using social media networks to connect to their customers and conduct their business online [2, 3, 4]. Many transactions are carried out over the network connection. These transactions may involve large sums of money being transferred between businesses or smaller sums between individuals. No matter how small these transactions are, they should be conducted in a safe environment. Implementing honeypots in a business environment may benefit the business in two ways. Firstly, it can provide some form of defence to the real system. The attacker may be diverted away from the actual system into the honeypot. Secondly, the lessons learned from the attacks can be used to build even better defence mechanisms.

This paper presents a survey of the emergent trends with respect to honeypot research. The aim is to identify trends and opportunities for guiding the work of honeypots researchers in the field.

The rest of the paper is laid out as follows: Section II gives a brief background on the honeypot concept. The methodology adopted by this study is explained in Section III. This is followed by the research findings in Section IV. Sections V and VI contain a discussion of the findings and the conclusion, respectively.

II. BACKGROUND

Spitzner [5] details a timeline from 1990 through to 2002 on network security tools. This timeline indicates that as access to the internet grew, the need for tools that could be configured easily to monitor the network also increased. According to Spitzner [5], Fred Cohen's Deception Toolkit—which was used to publish fake services that could be attacked by individuals intending to break down the system—was the first example of a honeypot tool. Researchers have different definitions of what a honeypot is. Pelletier and Kabay [6] define a honeypot as a system that serves the purpose of being exploited in a manner contrary to the terms of services of that system. According to Azadegan and McKenna [7], the purpose

of a honeypot is to determine tools and techniques used by hackers in order to prevent these hackers from gaining access to the real system. Joshi and Sardana [8] use the concept of dynamic honeypot to refer to a honeypot that learns from the type of attack in order to keep itself relevant. In this way, when the attacker discovers the weakness in the honeypot design, the honeypot reconfigures itself and adapts to the situation. Mokube and Adams [9] define the term "honeypot" to characterise a computer system implemented as a decoy network that is used to entice possible attackers into exploiting the system with various tools within their hacking toolkit. As mentioned in Section I, this paper focuses on honeypots as a defence strategy for monitoring attack behaviours and keeping a record of those suspicious actions.

Honeypots can be categorised into four types: shadow honeypots, honeynets, honeyfarms and honeytokens. Anagnostakis et al. [10] define a shadow honeypot as a mixture of a honeypot and anomaly detection—any behaviour on the network that is considered abnormal or suspicious according to a set of predefined rules or signatures. If such signature is found, then the system is alerted of anomalous behaviour and the suspicious network traffic is diverted to a shadow network. Should this be a hostile error, the traffic is redirected to the real system, as mentioned by Briffaut et al. [11]. Spitzner [5] and Hoepers et al. [12] view honeynets as research-oriented honeypots which are aimed at gathering information about the attacker. The attacker's toolsets and methods can be collected as such.

Another form of honeynet is a honeyfarm. Jiang et al. [13] describe honeyfarms as a dedicated set of honeynets. There are few semantic differences between a honeynet and a honeyfarm, according to Jiang et al. [13].

Honeypots can be characterised by their interaction level, deployment modes and deployment categories. These characteristics are discussed in the following subsections.

A. *Interaction level*

The interaction level of a honeypot is the degree to which an attacker has access to the system services. A honeypot can have a low, medium or high interaction level, depending on the functionality and availability of system services, according to Gibbens and Rajendram [14]. Bhumika and Sharma [15] describe a low interaction honeypot as one where an interaction with the system service is kept to a minimum. This means that the honeypot cannot be used to launch attacks to an external system, which is the advantage of this interaction level. However, attackers can easily discover the true nature of this system as a honeypot.

Similar to a honeypot with a low interaction level, a honeypot medium interaction level implements a reduced set of services. The services at this level do, however, have some intelligence built into them so that they are not immediately recognisable as honeypots.

High interaction honeypots implement a full operating system with the entire range of system functionality available to the attacker. This means that the honeypot can be used to launch attacks on the actual system, in contrast with low and medium interaction levels. Honeypots implementing this level of interaction hide their true nature from their attackers better than low and medium interaction level implementations.

B. *Deployment* modes

Scottberg et al. [16] explain that honeypots can be deployed in one of three deployment modes. The modes are deception, intimidation or reconnaissance.

The deception mode manipulates the hacker into believing that the responses being received are from the real system. Martin [17] refers to the deception mode as a system which is used as a decoy and contains security weaknesses in order to attract hackers. Deception techniques ensure that attackers utilise every hacking tool at their disposal so that the honeypot can gather as much information as possible about the attack, which can be used to protect the real system. According to Scottberg et al. [16], a honeypot is involved in deception activities if its responses can deceive an attacker into thinking that the responses are from the real system.

In intimidation mode, the attacker is made aware of the measures in place to ensure the security of the system. A warning may be issued to notify the attacker that the system is being monitored and actions of the attacker are recorded. This message may scare off some of the attackers and leave only experienced hackers with in-depth knowledge and hacking techniques. In this way the possibility of new hacking tools and techniques being recorded by the honeypot is high. This may result in a honeypot recording high-quality information that can be used to reinforce the security measures on the real system.

One of the most important features of a honeypot is its ability to capture and record new attacks on the system. In reconnaissance mode, the honeypot is used to determine the tools and techniques used by the attacker. This information is used to implement heuristics-based rules that can be applied in intrusion detection and prevention systems. In this mode, the honeypot is used to detect both internal and external attackers of the system. Internal attackers are attackers from inside the organisation whilst the external attackers are outsiders.

The difference between the reconnaissance and intimidation deployment modes is that whilst both modes attempt to extract usable data from incursion, an intimidation deployment mode will actively try to keep the attacker on the system as long as possible by feeding back information which is designed to be engaging. This may result in further revealing other networks within the organisation, which could be additional honeypots.

C. *Deployment* categories

Spitzner [5] classifies honeypots as being used mainly in production and research environments. Production honeypots protect systems that are being utilised by an organisation. They are used in cases where there is an active threat to a system. According to Spitzner [5], this honeypot is used to catch internal attackers of the system.

Research honeypots, on the other hand, are implemented to track and record behaviour rather than to catch and punish intruders. They are set up to determine new methods of attacks

and to translate this information to new defensive strategies to combat hackers.

III. METHODOLOGY

The research methodology adopted in this study was a systematic review of extant literature in honeypot research. This involved searching for data records in the field involving network security terminology.

The study collected quantitative data from various sources including books, journal articles, conference proceedings and the Web using web search engines such as Google. The data rendered by this search engine was categorised accordingly by source type. The following attributes were identified and recorded for each source: type of publication, year of publication, country of origin and subject theme of a publication.

Within the publication type attribute, the source could be one of the following: book, Government Gazette, journal article, conference paper, or a report. The year attribute was within the range of 1990 to 2013. The keywords that were used to identify these sources were network security, honeypots, honeyfarms, honeynets and intrusion detection systems.

After collecting this data, the sources were reviewed to determine if they were relevant and the duplicated data was removed. During the collection of the data, it was discovered that some sources contained more than one theme. In such cases the dominant subject theme was taken as the main theme of that source.

IV. RESEARCH FINDINGS

This section presents the findings of the study. The values produced on the y-axis are actual totals and should not be interpreted as percentages.

A. Research material by year

Fig. 1 shows a view of the year of publication for the books, articles and web pages that were found by this study. As depicted, the bulk of the research around honeypots was conducted in the 2002—2003 timeframe. Possible reasons for the interest around this timeframe may have to do with the general increase in Internet usage and the lack of Internet security standards at that time. This meant that the security over the Internet became important to its users and the people encouraging its use.

A possible scenario causing this high rise could be that those latecomers to the Internet who had waited for Y2K to come and go before adopting technology were beginning to be concerned about their systems being online. Another possible scenario could be that the generation of children who had grown up with a PC as a more commonplace household item were now starting to buy their own equipment and exploring the World Wide Web, which now had many more connections made to it by companies than before.

Another peak is seen from 2006 up to 2012. Initially, the Internet was used mainly to perform business transactions and usually on desktop devices. The drop down of computer size to mobile devices increased the number of devices that are connected to the Internet. The Internet is no longer used only to conduct business transactions, but it is also a social networking platform where people may interact on a social level. This is one of the reasons that during this period, network security experts were concerned about the security in this type of network. The last observation worth mentioning about Fig. 1 is a seeming decrease in publications in 2013. Much of the research for this paper was done during 2013, which may be the reason for the bar appearing to have dropped in 2013. One should keep in mind that the studies done in 2013 were not available at the time. However, this does not mean that the security concerns decreased.

B. Research material by country of origin

Fig. 2 shows the country of origin of the research publications used for this paper. The publications come from seven countries and one small group of countries referred to as "International group" in the figure. The seven countries are Australia, Brazil, India, France, United States of America (USA), Germany and South Africa. The international group contains a number of countries that produce the lowest publications per country. This grouping makes the graph in Fig. 2 easy to read.

The USA has produced the majority of publications, as can be seen in Fig. 2. Whilst this may seem to suggest that Americans are either more concerned with security, or are victims of hacking more often than the other countries' citizens, no valid conclusion could be drawn in this regard. Moreover, the reliance on searches from Google Scholar with English language results may have an influence on these results. This is evident when one considers countries such as India, which does not use English as its language of choice when it comes to academic writing. Even so, one should bear in mind that the USA is a developed country with probably more businesses and people relying on network connections for their

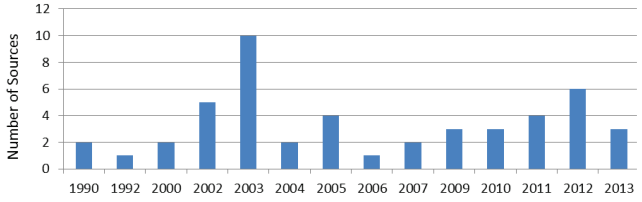

Figure 1. References by year

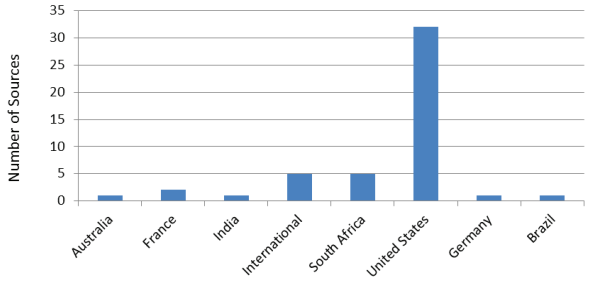

Figure 2. References by country of origin

day-to-day activities in comparison with developing countries. In such situations, the risk of network threats is likely to be high, thus requiring attention.

The USA is followed by South Africa, whose academic writing is also in English. Looking at the figure, it can be seen that South Africa and the international group have the same number of publications. Before conclusions are drawn, one should bear in mind that the international group is not a single country and therefore it would be not appropriate to compare it with single countries. The reason for including the international group in the figure was to acknowledge the existence of those countries for their contribution to the knowledge. In any case, this contribution has already been recorded in other comparisons, for example: in publications by year.

C. *Research material by publication type*

During the search for sources of information for this study, the majority of the relevant information was sourced from the Web and journal articles as shown in Fig. 3. Other forms of resources were utilised as indicated in the figure. These sources are books, conference papers, journal articles, the Government Gazette and web information.

It is interesting to note that the number of web publications tops the list, followed by journal articles. It is easier to publish on the Web than in any other format. This may explain why web publications exceed the other forms of publications. Another reason for this may be the ease of accessing such publications.

Another factor to note is the number of conference papers compared to journal articles. In academia, researchers prefer to publish in a journal. The final destination of some of the conference papers is in a journal publication. This may be due to academic incentives that come with a journal publication. Source from the Government Gazette are the least in the list. This publication platform is not likely to be used by academics and researchers to publish their research.

D. *Research material by subject*

In this section every information source used in the research was categorised within broad subjects addressing the main audience of the publication. Several subject themes related to network security were identified, namely ethics, security, honeypots, intrusion detection systems, tools et cetera, as shown in Fig. 4.

The bulk of the research was based on the honeypot and its various derivatives, including honeynets, shadow honeypots and honeytokens. Whilst alternative ethics papers were reviewed, many were discarded due to duplication. Fig. 4 shows the results of this categorisation.

In a survey on future trends in honeypot research, five areas were identified by Bringer et al. [18]. These areas include the development of new types of honeypots to cope with new security threats, the use of honeypots output data to improve the accuracy in detecting threats, honeypot configurations lowering false positives, counteracting honeypot detection by intruders, and ethical issues related to the use of honeypots.

V. DISCUSSION OF FINDINGS

The results of this study suggest that the honeypot research has been rising since 2006 (Fig. 1). It was reasoned that this is because the number of people who use devices that are connected to the network has been rising since then. The use of smart phones has enabled people to be connected to the network without even being aware of this connection. The number of companies conducting their business over the Internet or using social media is also increasing. As a result, the security of these connections cannot be ignored.

The trustworthiness of this study was supported by the types of publication sources that were consulted as seen in Fig. 3. One would agree that academic papers and books are more reliable sources of information in such investigations. Thus, the credibility of these primary source types is unquestionable, making the observation that honeypot research is on the rise justifiable.

The findings have indicated that researchers are showing interest in the honeypots as depicted in Fig. 4. This includes honeypot derivatives—honeynets, honeyfarms, etc. One should remember that apart from being used to catch intruders, a honeypot can also be used as a research tool, as mentioned in II.C. Research honeypots may yield knowledge that can be used to develop new tools and techniques to improve network security. In some cases where such tools exist, they can be improved using such knowledge. This means that there are still

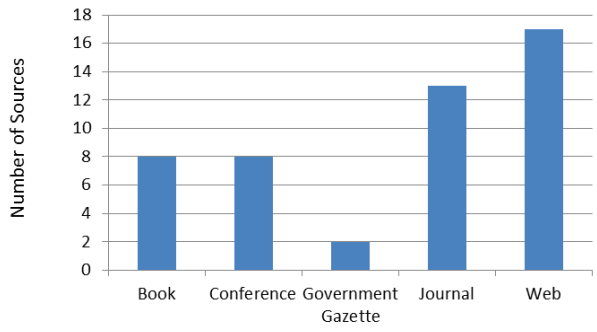

Figure 3. References by publication type

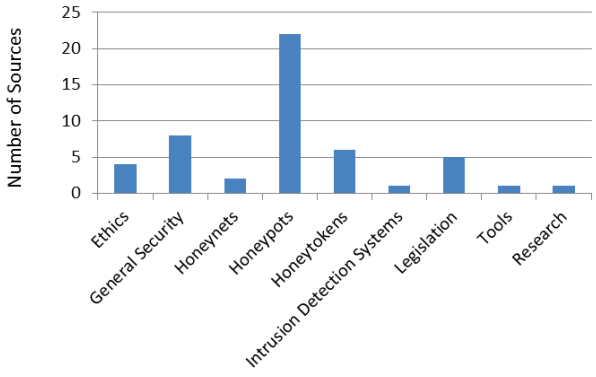

Figure 4. References by subject

opportunities for the researchers in the field to further investigate the use of honeypots, as mentioned in the previous section. As mentioned in IV.B, the research in this study was limited to publications in English and the majority of searches were done using Google Scholar or the internal search engines contained in the IEEE Xplore and ACM Digital Library systems. Therefore this research may be missing critical views in non-English publications. Nevertheless, this investigation has demonstrated which of those academically English-oriented countries have been involved in network security studies. According to Fig. 2, the USA leads in network security. The information in Fig. 2 indicates which countries may have a better understanding of the honeypots defence mechanism. On the other hand, the question could be asked whether countries with fewer publications have better methods of dealing with these threats if they have less to say about them. The data collected during the study was placed in four categories: year of publication, country of origin, source type and subject. Future studies could consider integrating these categories instead of examining them independently. This would allow a clearer understanding of, for example, which country is showing concern and in what year. That would explain why some countries have fewer publications than others. The possibility is that they have learned a lot from countries with more publications and have merely implemented the solutions of those countries. The areas of future research mentioned by Bringer et al. [18] present various opportunities for researchers. Researchers may be interested in developing new honeypots, using the data obtained from research honeypots to improve existing honeypots or develop new ones. The configuration of those honeypots must be done in such a way that they do not bore the attackers or scare them off. It is important that attackers are kept fully engaged in order to learn as much as possible about their actions. The legal and ethical issues are also an important area of concern regarding the use of honeypots. Other researchers believe that it is unethical to use honeypots to fool attackers without their knowledge.

VI. CONCLUSION

The purpose of this study was to investigate emergent trends concerning honeypot research. The study involved studying various sources available in the literature on honeypots as a network security measure. The literature reviewed indicates that honeypots are of interest when it comes to network security. The amount of information that is produced by countries like the USA indicates that honeypots have a role in network security. Honeypots can be used to catch network intruders and also to learn the techniques used by these intruders to gain access to network systems. This knowledge can be used to improve other honeypots or the system itself. In addition, new systems other than honeypots may be developed in future using this knowledge.

REFERENCES

[1] E. Kaspersky, "Kasperskty lab africa," 2010, [Online]. Available: http://www.kaspersky.co.za/. (Access Date: 15 April, 2013).

[2] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," *Business Horizons*, vol. 53, no. 1, pp. 59 – 68, 2010.

[3] R. V. Kozinets, "The field behind the screen: using netnography for marketing research in online communities," *Journal of marketing research*, vol. 39, no. 1, pp. 61–72, 2002.

[4] A. M. Muniz Jr and T. C. O'guinn, "Brand community," *Journal of consumer research*, vol. 27, no. 4, pp. 412–432, 2001.

[5] L. Spitzner, *Honeypots: Tracking Hackers*. Addison Wesley, 2002.

[6] R. Pelletier and M. E. Kabay, "Honeypots, Part 2," 2003. [Online]. Available:http://www.networkworld.com/newsletters/2003/0512sec2.html, (Access Date: 15 March, 2013).

[7] S. Azadegan and V. McKenna, "Use of honeynets in computer security education," in *Computer and Information Science, 2005. Fourth Annual ACIS International Conference on*, 2005, pp. 320–325.

[8] R. C. Joshi and A. Sardana, *Honeypots: A New Paradigm to Information Security*. Enfield: Science Publishers, 2011.

[9] I. Mokube and M. Adams, "Honeypots:Concepts, Approaches and Challenges," in *Proceedings of the 45th annual SouthEast Regional Conference of the ACM*, vol. 4, Winston-Salem, 2005, pp. 321–326.

[10] K. G. Anagnostakis, S. Sidiroglou, P. Akritidis, M. Polychronakis, A. D. Keromytis, and E. P. Markatos, "Shadow Honeypots," *International Journal of Computer and Network Security*, vol. 2, no. 9, pp. 1–16, 2010.

[11] J. Briffaut, J.-F. Lalande, and C. Toinard, "Security and results of a largescale high-interaction honeypot," *Journal of Computers 4.5 (2009)*, pp. 395–404, 2009.

[12] C. Hoepers, K. Steding-Jessen, and A. Montes, "Honeynets Applied to the CSIRT Scenario," in *Proceedings of the 15th Annual Computer Security Incident Handling Conference*, 2003.

[13] X. Jiang, D. Xu, and Y.-M. Wang, "Collapsar: a vm-based honeyfarm and reverse honeyfarm architecture for network attack capture and detention," *Journal of Parallel and Distributed Computing*, vol. 66, no. 9, pp. 1165–1180, 2006.

[14] M. Gibbens and H. V. Rajendram, "Computer Science Department," 2012. [Online]. Available: http://www.cs.arizona.edu/ collberg/Teaching/466-566/2012/Resources/, (Access Date: 20 April, 2013).

[15] L. Bhumika and V. Sharma, "Use of Honeypots to Increase Awareness regarding Network Security," *International Journal of Recent Technology and Engineering*, vol. 1, no. 2, pp. 171–175, 2012.

[16] B. Scottberg, W. Yurcik, and D. Doss, "Internet honeypots: Protection or entrapment?" *Technology and Society International Symposium, (ISTAS'02)*, pp. 387–391, 2002.

[17] W. W. Martin, "Honeypots and Honeynets: Security through deception," 2001. [Online]. Available: http://www.sans.org/reading-room/ whitepapers/attacking/honeypots-honey-nets-security-deception-41,(Access Date: 24 March, 2013).

[18] M. L. Bringer, C. A. Chelmecki, and H. Fujinoki, "A survey: Recent advances and future trends in honeypot research," *I. J. Computer Network and Information Security*, vol. 10, pp. 63–75, 2012.