Research trends in digital forensic science: An empirical analysis of published research

Ibrahim Baggili, Afrah BaAbdallah, Deena Al-Safi, Andrew Marrington

Zayed University, Advanced Cyber Forensics Research Laboratory
Abu Dhabi, United Arab Emirates, P.O. Box 4783
{Ibrahim.Baggili,Andrew.Marrington}@zu.ac.ae

Abstract. Digital forensic science is a new discipline. In order to advance and improve this science, stakeholders should stay abreast over the research trends in this domain. This research studied, categorized and analyzed a sample of five-hundred publications (n=500) from this discipline. The results indicated that the rate of publication in this domain continues to increase over time. Additionally, results showed an overall lack of anti-forensics research where only 2% of the sampled papers dealt with anti-forensics. In terms of research methodology, the results indicated that 17% of the sampled publications were secondary research, 36% were exploratory studies, 33% were constructive and 31% were empirical. The results also indicated a lack of basic research in this scientific discipline where most of the research (81%) was applied, and that only 19% of the sample was categorized as basic research. Additionally, results exemplified a lack of quantitative research in the discipline, with only 20% of the research papers using quantitative methods, and 80% using qualitative methods. Furthermore, results showed that the largest portion of the research (42.9%) from the examined sample originated from the United States. The findings also showed a lack of cooperative research between academia and industry, where only 10% of the research studies examined where a collaborative effort between industry and academia. Lastly, the findings indicated an increase in the disparity between the number of published articles and the number of cited articles over the years possibly indicating isolation amongst researchers in this domain.

Keywords: Digital forensic science, research trends, research methodologies, challenges in digital forensics science.

1 Introduction

Cybercrime initially emerged as a threat to computer users and businesses; it now impacts entire nations. Internet usage continues to rise and so does this threat [1]. Yet, most computer users remain unconscious of the drastic impact it has on their daily lives. The statement "The Internet is the crime scene of the 21st century" as written in the Wall Street Journal, is a realistic indicator of the current times [2].

Rogers and Seigfried in 2004 reported that cybercrime is constantly on the rise,

spurring a massive progress in digital forensic science (DFS) [3]. This has consequently lured the attention of scientists towards a subset of DFS – computer forensics, establishing it as a recognized scientific discipline [4][5].

Patzakis in 2003 described computer forensics as a process of collecting, preserving, analyzing, and presenting electronic evidence where a computer has been an instrument to committing a crime [6]. This investigative methodology is used to reconstruct computer evidence as well as examine digital media storage devices in order to find electronic evidence which could lead to the source of the crime and its perpetrator(s). Furthermore, computer forensics is recommended whenever the security of an organization or company has been breached. In such a scenario, system administrators begin investigations by acquiring and analyzing the collected digital evidence.

Research has been conducted and articles published discussing various topics in DFS. Some researchers have illustrated specific definitions and processes in digital forensics [7], whereas others have published studies addressing anti-forensics [8]. Additionally, certain researchers have focused their attention to incident response and best practices when a computer crime occurs [9]. It is beyond this research paper's scope to provide a complete overview of all the research conducted under the DFS umbrella. Nonetheless, it is critical for scientists as well as practitioners to keep up with research trends associated with the science of digital forensics to acknowledge and further investigate gaps in the domain.

This research provides a strong primary contribution to this new scientific discipline, as it empirically studies research trends in the field. The primary goal is to empirically explore the path that DFS is moving towards through the categorization and analysis of a sample of five-hundred (n=500) publications issued between 1992 and 2011.

2 Literature Review

DFS is at its infancy and continues to be of utmost importance. Governmental agencies are obliged to depend on the scientific and private communities to derive novel methods and tools that allow the extraction and preservation of digital evidence in a scientific and law-abiding manner. Given the importance of this field and its impact, it is essential to collect, analyze, and categorize research in this scientific domain. This can help shed light on the discipline, aiding in a more appropriate response to cybercrime while contributing to the development of the science and professional practice in this field.

Garfinkel in 2010 argued that there is a genuine need for a well defined and collaborative approach to be undertaken by the researchers and institutions in digital forensics [10]. Garfinkel stated that "Without a clear strategy for enabling research efforts that build upon one another, forensic research will fall behind the market, tools will become increasingly obsolete, and law enforcement, military and other users of computer forensics products will be unable to rely on the results of forensic analysis" [10].

In order to combat challenges mentioned in academic literature, it is critical to consistently and empirically study research trends in DFS under a framework where research in the discipline is collected, categorized, and analyzed. The results can aid researchers and practitioners in keeping abreast over the trends in the scientific domain, as well as ensuring that they are on target with any intended scientific goals.

The concept of research trends includes creating a trend map from research papers and patents and enabling the discipline's stakeholders to grasp the outline of technical trends in a particular field [11]. This concept is not new and has already been used in various disciplines such as Psychology [12], Biology [13], and Sociology. Furthermore, research trends guide the scientific community in solving challenges and potential obstacles that hinder the process of the discipline's development.

Some scientists have illustrated interest in DFS research trends reflected by their research on the future of the discipline. Rogers and Seigfried in 2004 disseminated a survey to study and characterize the top five issues in computer forensics. In their paper, they addressed the main challenges in the field, as well as issues pertaining to having a defined standardization and modular approach for data representation and forensic processing.

Moreover, in 2007, Chichao, Wenyuan, and Weiping [14] presented results in their study which aimed at exploring trends in computer crime and cybercrime research from 1974 to 2006. In their research, two-hundred and ninety two (n=292) papers on computer crime and cybercrime publications were drawn from the ISI Web of Science, the Science Citation Index (SCI), and the Social Science Citation Index (SSCI). Their results indicated that many papers were written in English, and most articles came from the U.S.A.

The purpose of this study was to explore the trends in DFS research from past till present. Publications for this analysis were drawn from scientific and professional publications such as Springer, Elsevier, Digital Forensics Research Conference (DFRWS), Journal of Digital Forensics Security and Law (JDFSL), National Institute of Standards and Technology (NIST), Small Scale Digital Device Forensics Journal, International Journal of Digital Evidence (IJDE), Journal of Digital Forensic Practice, International Journal of Electronic Security and Digital Forensics (IJESDF).

What made this research study unique is that the researchers did not disseminate a survey; rather, they studied, categorized and analyzed the existing literature in DFS to extrapolate an overview of the scientific discipline and the research trends associated with it over the years.

3 Methodology

The procedures followed during the data collection phase were empirical. First, the authors depended on credible publication venues to collect a sample of publications. The International Journal of Digital Evidence, Digital Forensics Research Conference, and Springer and Elsevier were powerful resources for collecting the data needed for the study. Using the collected articles, the authors built a database containing a sample of five hundred (n=500) research papers related to DFS. The breadth in the

publications helped cover a wide range of research topics in the discipline across different time periods.

The process of categorizing the data spanned over two months. It started in the middle of June 2011 and carried on until the middle of August 2011. Here, the authors note that the categorization process was manual. Because the process was manual, bias could have possibly been introduced into the methodology due to human error. The authors note that this is a limitation in this study and that the researchers strived to remain accurate throughout the categorization phase.

During the categorization of the papers, each paper that was added to the database was examined and classified using the following categories:

- Publication year
- Forensic type (Forensic/Anti-Forensic)
- Research type (Primary, Secondary)
- Research methodology type (Exploratory/Constructive/Empirical)
- Research category (Basic/Applied)
- Research method (Qualitative/Quantitative)
- Location/Country of the research
- The originator of the research (Academic/Business or company/ Co-operation of both)
- Cited/Not cited

Based on the abovementioned categories, the authors objectively classified each paper and documented that categorization accordingly.

4 Findings and analysis

The final database contained a sample of five hundred publications (n=500). The data for each classification category was then analyzed and graphs were created to extract general trends. The findings for each of the categories are shown in the sections that follow.

4.1 Publication year

In this category, the authors examined the percentage of publications produced over time, as shown in Figure 1. Figure 1 illustrates how the number of research publications increased over the years. Starting in 1992, the number of published papers was insignificant compared to the number of papers that were published in 2010. This trend indicates that the number of studies in DFS has steadily increased throughout the years, though there was a slight decrease in the number of published research between 2002 and 2003 then a steady output of publications between 2007 and 2008.

The research findings also highlight a dramatic decrease in the number of research publications published between 2010 and 2011. A reason for this drop could be that data collection ceased before the end of August 2011 and that publication houses typically issue papers that were presented in 2011 in 2012 editions of journals or

conference proceedings. Moreover, it is important to take into consideration the general length of time required to submit, accept, approve, and publish peer-reviewed research papers. The authors note that this is a potential limitation in the sample of papers collected. Hence, a prediction can be made that the annual increase of papers will progress as DFS continues to capture the attention of more researchers and organizations.

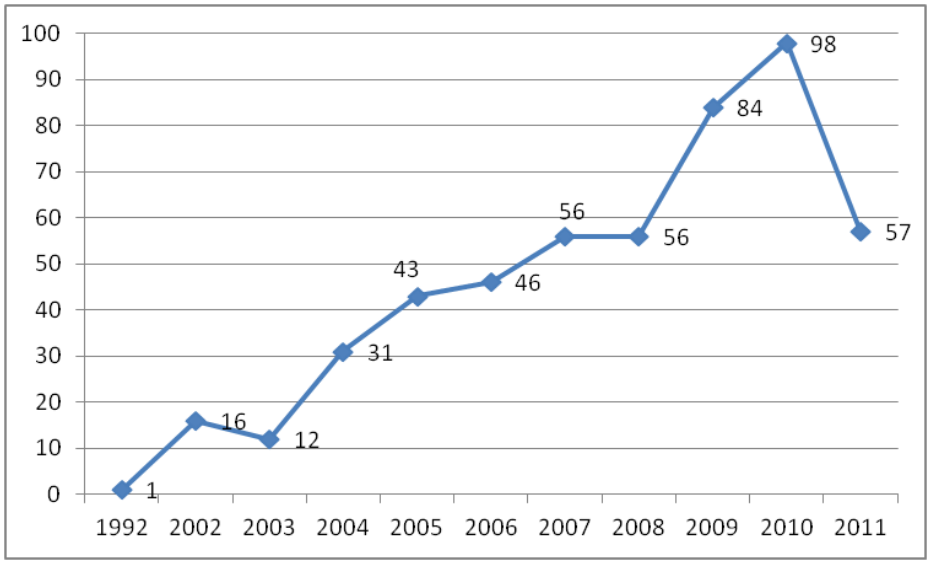

Fig. 1. Digital forensic science publications over time.

4.2 Forensic type

From the collected data, a conspicuous trend was noticed. After categorizing the research papers into forensic and anti-forensic related research papers, the results showed that only 2% of the studies discussed anti-forensics, while 98% of the publications discussed forensics.

One plausible explanation for this is that most scientific research aims at improving the effectiveness of forensic examination, whereas anti-forensics has the opposite focus. It is likely that much of the anti-forensics innovation occurs outside of the academic community altogether. Consequently, a relatively low proportion of development in anti-forensics appears in the peer-reviewed scientific literature. Irrespective of the reasons, what can be observed is that, overall, anti-forensics is neglected as a research topic in DFS.

4.3 Research type

The results in this category illustrate that 83% of the analyzed publications were categorized as primary research studies, and 17% were secondary research studies.

These results signify that DFS is being driven by primary research, reflecting the novelty and infancy of this science.

Another pattern the authors analyzed was research output type over time. Figure 2 illustrates the results obtained from that analysis.

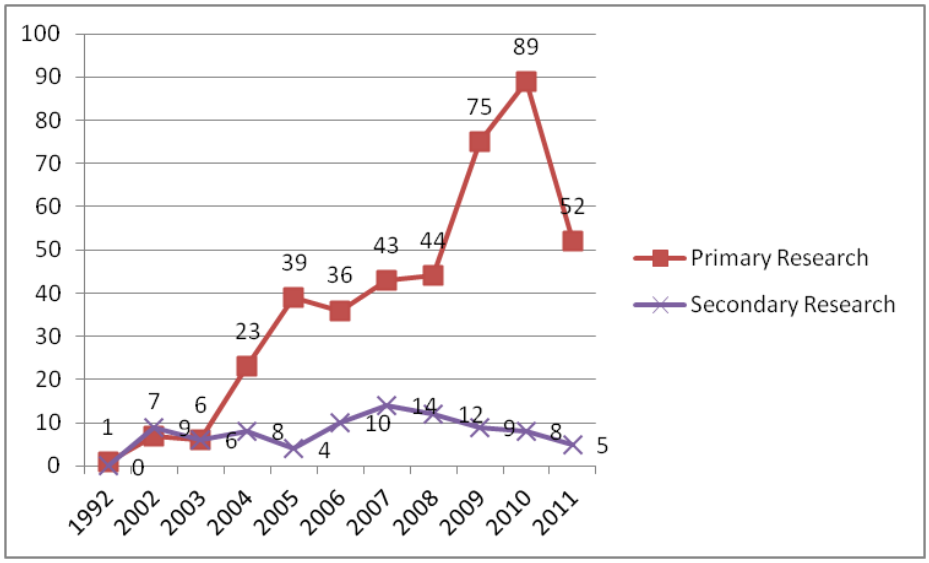

Fig. 2. Primary and secondary research over time.

From Figure 3, one can clearly observe that primary research studies continued to increase from 1992 until 2010 with some minor fluctuations. In 2010 it reached the peak with 89 publications. Again, a plausible explanation for the decrease from 89 to 52 publications in 2011 is that the publication sample used in this research was not representative of all the research studies that were conducted in 2011.

As for secondary research studies, Figure 2 demonstrates fluctuations in this type of research. Overall, there has been a slight decline in secondary research in recent years.

4.4 Research methodology

The authors examined the research methodologies used in each of the sampled publications. The three types of research methodologies used in the categorization process were: constructive, empirical and exploratory. Table 1 shows the results obtained from the data analysis process.

As shown in Table 1, the largest percentage of papers, 36%, used an empirical methodology. The reason for this may be due to the fact that this field is still new. Therefore, additional knowledge can be gained by using methods such as direct and indirect observation or experience. Furthermore, a reason for the high percentage of utilization of the empirical research methodology could be linked to the high

percentage of primary research, which depends on collecting original data after gaining knowledge from a direct observation [15].

Table 1. Research methodologies used.

Research methodology	% of sample papers
Empirical	36 %
Exploratory	33 %
Constructive	31 %

The results also indicate that 33% of the research papers used an exploratory methodology. The plausible explanation for that could be that since DFS is new, many exploratory studies are being pursued to gain a deeper understanding of the science.

Finally, 31% of the research papers used a constructive research methodology. Constructive research is highly linked to the computer sciences, and so is DFS. A plausible explanation for this finding is that most researchers in this domain come from a computer science background, thus many are trained to use constructive research methodologies. To examine the research methodology over time, data was analyzed further, as shown in Figure 3.

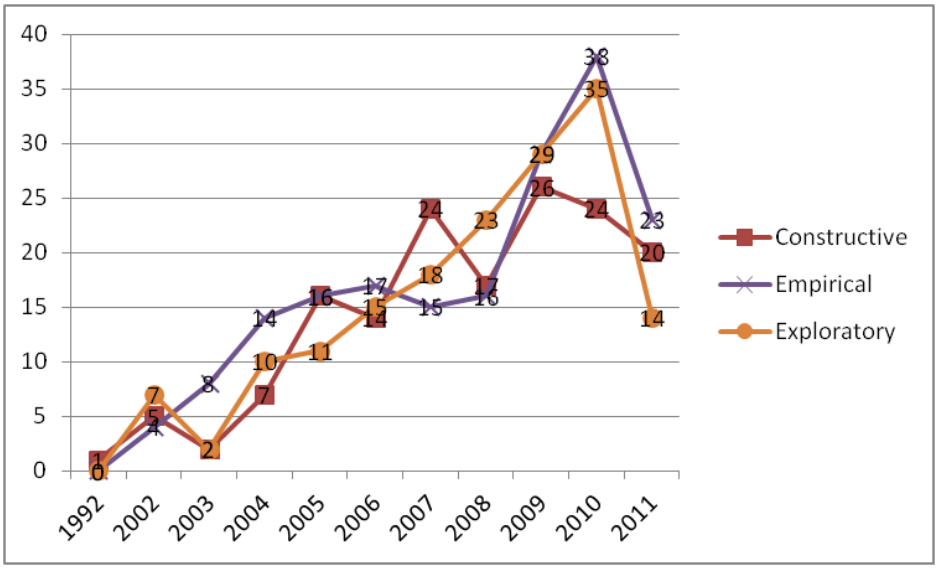

Fig. 3. Research methodologies over time.

Speaking generally, there has been consistent growth in the number of exploratory studies, with some fluctuation in recent years in the rates of publication of constructive and empirical studies. The authors note that the figures for 2011 are inaccurate because of the sample's misrepresentation of the literature that was

published in 2011.

4.5 Research category

Research can be categorized as applied or basic research. The collected data demonstrates that 81% of research studies are applied, and only 19% are basic. This is shown in Table 1.

Table 2. Research category (Applied and Basic).

Research category	% of sample papers
Applied	81 %
Basic	19 %

Applied research deals with solving practical problems and generally employs empirical methodologies [16]. This sustains the previous trends described. In contrast, basic research tends to expand the knowledge and understanding of essential principles that might not add any direct benefit or conclusions. This might be a cause of worry to the field of DFS, since it is new. The authors speculate on whether or not more basic research should be pursued by scientists in this domain to strengthen the foundational elements of this discipline.

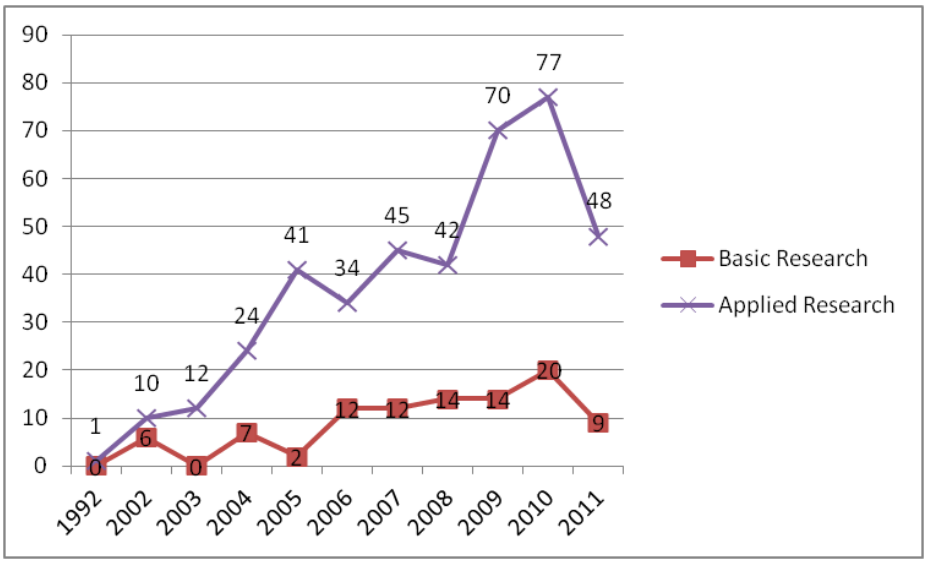

Fig. 4. Research category (Applied and Basic) over time.

To gain a more thorough understanding, the authors analyzed the category of research over time as shown in Figure 4. The gap between the number of applied research publications and basic research publications is presently wide (57 in 2010, 56

in 2009). From year 2006 to 2011, the rate of basic research publication was reasonably steady (with a slight increase in 2010 followed by a decrease in 2011).

The authors speculate that these results could be attributed to the fact that DFS is a discipline that has been driven by experienced and applied practitioners in the field that may not have had traditional academic research training. Many stakeholders in DFS argue that the nature of the field is applied and therefore the amount of applied research in this domain reflects that notion.

4.6 Research method

Table 3 shows another significant finding in DFS research trends. It illustrates that 80% of the research studies were categorized as qualitative and only 20% were quantitative. The fact that the qualitative method investigates the why and how of decision making makes it understandable as to why there is such a high percentage of qualitative research [16]. In order to further investigate the research methods used, the authors analyzed the research method used over time as shown in Figure 5.

Table 3. Research methods used.

Research method	% of sample papers
Quantitative Research	80 %
Qualitative Research	20 %

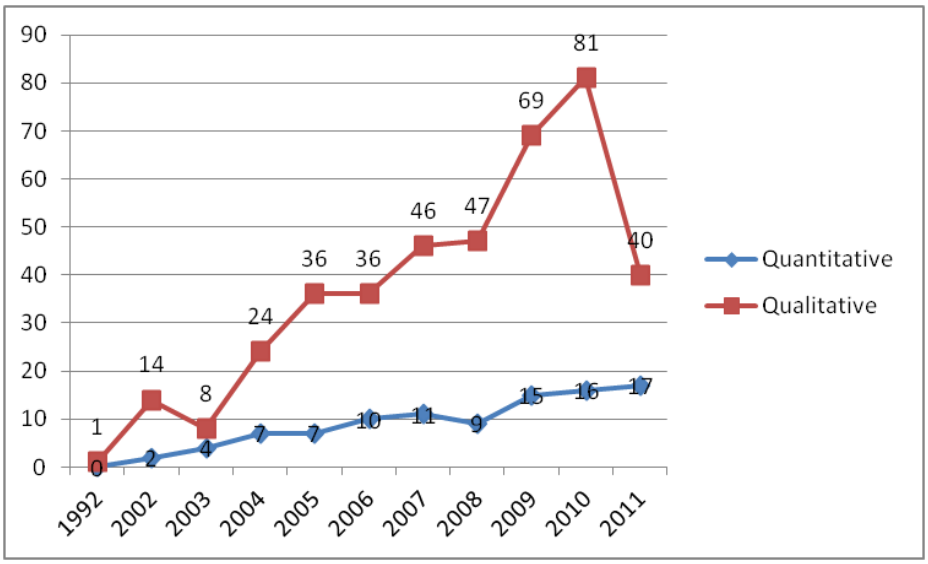

Fig. 5. Quantitative and qualitative research over time.

Figure 5 clarifies the relationship between the research methods and time. The volume of quantitative research (as measured by published research papers) has

steadily increased over the years. Qualitative research, however, experienced two sharp spikes, one in 2004 (24 from 8 in 2003), and in 2009 (69 from 2008, and 81 in 2010).

Overall, these graphs illustrate that both research methods are increasing. The significant drop in qualitative research in 2011 could also be attributed to the aforementioned sample problem; the data was collected before all the 2011 research studies were published.

4.7 Location (country of origin) of research

It is important to highlight the location of research publications because it leads to the discovery of the countries that are pursuing research initiatives in DFS. Therefore, the data collected was classified based on the institution and/or organization's country that issued the study. Some of the publications were issued in one country, yet a few were issued in co-operation between international universities and communities. Table 5 shows each country and the number of published articles released from that specific country.

Table 4. Publications by country.

Country	# of Papers	% of sample
USA	228	42.9
UK	49	9.2
Australia	37	7.0
China	23	4.3
Korea	22	4.1
India	17	3.2
Germany	16	3.0
Ireland	15	2.8
Italy	13	2.4
Taiwan	10	1.9
Canada	9	1.7
France	7	1.3
Japan	7	1.3
Malaysia	7	1.3
Hong Kong	6	1.1
UAE	6	1.1
Netherlands	6	1.1
Singapore	6	1.1
Sweden	5	0.9
Norway	5	0.9
South Africa	5	0.9
Belgium	3	0.6
New Zealand	3	0.6
Poland	3	0.6
Greece	2	0.4
Brazil	2	0.4
Finland	2	0.4
Iran	2	0.4
Switzerland	2	0.4
Saudi Arabia	2	0.4
Turkey	2	0.4
Algeria	1	0.2

Croatia	1	0.2
Romania	1	0.2
Luxembourg	1	0.2
Indonesia	1	0.2
Mexico	1	0.2
Pakistan	1	0.2
Uganda	1	0.2
Spain	1	0.2
Qatar	1	0.2

Table 5 illustrates that the United States of America holds the highest number of publications at 42.9% of the total sample. The United Kingdom comes in second place with a significant difference in percentage at 9.2%. China, Korea, India, Germany, Ireland, Italy, Taiwan, and Canada follow with different percentage variations at 4.3%, 4.1%, 3.2%, 3.0%, 2.8%, 2.4%, 1.9%, and 1.7% respectively.

4.8 Research originator

One of the categories used in this study to classify research articles was the originator of the research studies. Some of the papers were published by professors and academic experts, whereas others were prepared by digital forensic practitioners. Additionally, some of the publications were a cooperative effort between academia and private sector organizations. There have been continuous deliberations amongst experts in DFS regarding a stronger collaboration between academia and private sector with regards to DFS research. The authors thought it would be interesting to explore how much of the research originated from academic institutions, how much originated from companies, and lastly, the amount of publications in which companies and academic institutions jointly collaborated on. Table 6 shows the percentage of papers categorized by the originator.

Table 5. Research originator.

Originator	% of sample papers
Academic	60 %
Industry	29 %
Joint	10 %
N/A	1 %

Table 6 depicts that 60% of the publications were issued by academics. This high percentage indicates that universities and academics are the most productive in terms of research in DFS. Furthermore, 29% of the publications were issued by companies or organizations that were either interested or invested in DFS. Lastly, only 10% of the research papers stemmed from a cooperative effort between academics and organizations. The authors couldn't trace the origin of 1% of the collected publications.

These results illustrate a clear dichotomy between academia and organizations when it comes to DFS research. The authors understand the importance of

collaboration between academics and private organizations since the science of digital forensics is new and concurrently in practice.

4.9 Cited papers

It is accepted practice to regard impact publications to have a significant number of citations. Generally a large number of citations for a publication indicates that it is useful, effective, and in demand. From the five-hundred (n=500) publications that comprised the study, some were cited, and some were not. During this process, Google Scholar was used in order to check if an article was cited or not. Table 7 shows the percentage of cited and non-cited papers.

Table 6. Cited and non-cited articles.

Cited/non-cited	% of sample papers
Cited articles	66 %
Non-cited articles	34 %

The percentage of citied articles was 66%, unlike the percentage of non-cited articles which was 34%. This may indicate that DFS is gaining more attention by academics and organizations. Of course, the number of cited papers will continue to increase. Perhaps the more interesting metric is the proportion of papers which are cited – as this may indicate the proportion of the literature which is relevant and useful to other authors (and perhaps, indirectly, to industry and law enforcement). The analysis of the number of cited publications over time is illustrated in Figure 6.

Figure 6 also compares the number of published articles to the number of articles cited by publication year. The data shows that the number of publications in the sample continues to increase. A trend can be noticed as a significant increase in disparity between the number of articles published and the number cited articles starting 2009-2011. The authors speculate that this is most likely because newer articles have not been sufficiently exposed to other researchers for further work to be built on top of them yet, although it may also indicate a decrease in the proportion of published articles which may be regarded as seminal to DFS.

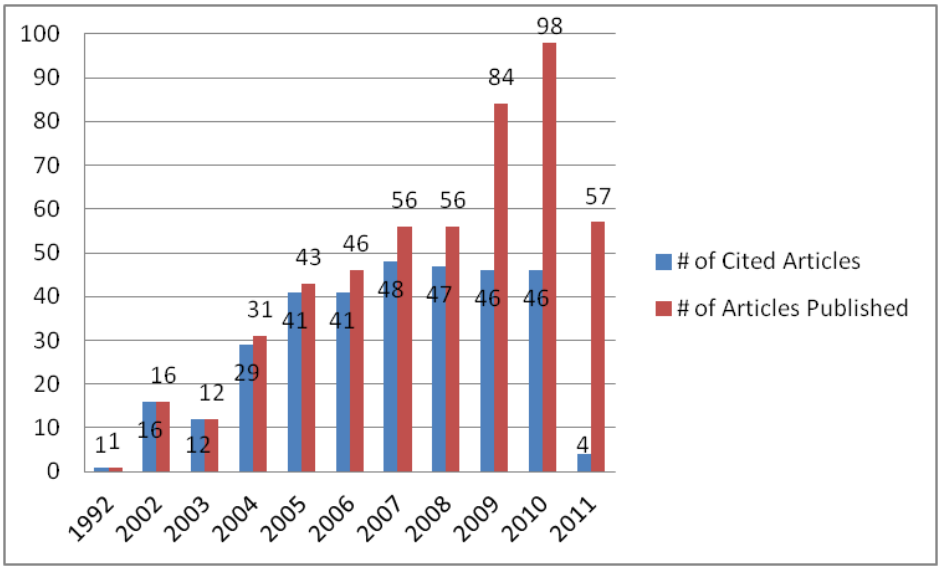

Fig. 6. Articles cited and published by year.

5 Conclusions and future work

DFS has captured the attention of the scientific community. This, in turn, has led to the increase in the number of studies and research in the discipline. Yet, it is still a new scientific discipline, which makes it difficult to clearly predict where this field is heading. Therefore, this research paper attempted to highlight the path that this new field of science is leaning towards; by collecting a random sample of papers (n=500) and classifying them into different categories in an attempt to arrive at research trend patterns.

There were a few limitations that affected the research in this paper. The main one was the difficulty in finding ontological research topics where research papers could be classified under. Due to this limitation, the authors were not able to classify the research publications into research topics such as media forensics, small scale device forensics, network forensics etc. This challenge also points to the idea that perhaps a more accepted ontology of DFS research topics should be researched and accepted within the scientific community.

The authors believe that the database of this primary research needs to be constantly updated for more accurate results, allowing DFS stakeholders to stay abreast over research trends across time. As the database grows, the sample size of the categorized research publications will increase as well. Constantly increasing the sample size will lead to better and more accurate future results.

This was the first empirical step in defining the path that this new discipline is taking. The authors hope that researchers will continue to expand on this research topic as the discipline of DFS continues to mature.

References

1. Wolf, S. (2004). A click away from a mugging. The Australian. Retrieved from https://intranet.stjohns.sa.edu.au/curriculum/infotech/ITissues/pdf/A_click_away_from_a_mugging.pdf
2. Solms, P. B. V. (2011). "S.Africa: The crime scene of the 21st century." Wall Street Journal.
3. Rogers, M. K. and K. Seigfried (2004). "The future of computer forensics: a needs analysis survey." Center for Education and Research in Information Assurance and Security, Purdue University, 656 Oval, West Lafayette, IN 47907, USA.
4. Whitcomb, C. M. (2002). "An Historical Perspective of Digital Evidence: A Forensic Scientist's View." International Journal of Digital Evidence Spring 2002 Volume 1, Issue 1.
5. Rogers, M. (2003). "The role of criminal profiling in the computer forensics process." Elsevier Ltd. Center for Education and Research in Information Assurance and Security (CERIAS), Purdue University.
6. Patzakis, J. (2003). Computer Forensics as an Integral Component of the Information Security Enterprise G. Software. California.
7. Oseles, L. (2001). "Computer Forensics: The Key to Solving the Crime." INSS 690.
8. Peron, C. S. J. and M. Legary (2005). "Digital Anti-Forensics:Emerging trends in data transformation techniques."
9. Al-Zarouni, M. and H. Al-Hajri (2007). "A Proof-of-Concept Project for Utilizing U3 Technology in Incident Response." School of Computer and Information Science, Edith Cowan University.
10. Garfinkel, S. L. (2010). "Digital forensics research: The next 10 years." a Naval Postgraduate School, Monterey, USA.
11. Nanba, H., T. Kondo, et al. (2010). Automatic creation of a technical trend map from research papers and patents. Proceedings of the 3rd international workshop on Patent information retrieval. Toronto, ON, Canada, ACM: 11-16.
12. Gauvin, L. and J. C. Spence (1995). "Psychology research on exercise and fitness: Current research trends and future challenges.
13. Marzluff, J. M., R. Bowman, et al. (2001). "A historical perspective on urban bird research: trends, terms, and approaches." Avian Ecology and Conservation in an Urbanizing World. Boston: Kluwer Academic.
14. Lu, C., W. Jen, et al. (2007). Trends in Computer Crime and Cybercrime Research During the Period 1974-2006: A Bibliometric Approach. Intelligence and Security Informatics. C. Yang, D. Zeng, M. Chauet al, Springer Berlin / Heidelberg. 4430: 244-250.
15. Teller, P. (2001). "Whither Constructive Empiricism?" Philosophical Studies 106(1-2): 123-123-150.
16. Gruman, J. C. (2003). "Basic vs. Applied Research: Finding a Balance." The Chronicle of Higher Education 49(29): B.20-B20.