Writing Scientific Papers

Antoni Martínez Ballesté

Introduction

During centuries, the advances of scientific research were out of reach for almost everybody. From the first civilisations to some centuries ago, the vast majority of people were more interested in surviving or in earning some nourishment to feed their families. The ancient Christian Church almost monopolised every field of knowledge. Sorcery was the keeper of the secrets of Medicine… and scientists, inventors and philosophers discussed their advances with rhetorical speaks in front of their disciples. That was in case of being lucky: a plethora of inventions were buried with their discoverers.

It was not until the XVII century that scientific knowledge started to be the basis of higher education. Naturally, only a few were able to attend lectures from renowned scientists and inventors. In that sense, scientific societies started to write down lectures and discussions so as to provide colleagues and students from other scientific institutions with a source of knowledge. Scientific texts were not only used for knowledge diffusion, but also for allowing other scientists to repeat experiments to improve or just assess the described results.

At the beginning of scientific publishing, scientists sent their works in a manuscript form. Currently, the use of software to publish scientific writings is mandatory. There are also thousands of institutions which produce thousands of scientific papers, which are selected and published by a large number of scientific journals. Moreover, the Internet is used by publishers and scientists themselves as a repository: scientific papers can be found in publishers' websites, or in scientists' home pages. This results in a huge quantity of scientific literature being produced day by day, which is as far as two mouse clicks from any scientist in the world.

This module is devoted to scientific writing in Engineering disciplines (for example, Computer Engineering, Electronics, etc.). In the first section, we describe the types of scientific papers. The main section of this chapter elaborates on the writing of each part of the scientific paper. The chapter concludes with a section containing some hints on language and style.

Goals

The goals to be achieved by studying this module are the following:

- Know the different kinds of scientific texts.
- Understand the structure of a scientific paper.
- Elaborate paper titles.
- Elaborate paper abstracts.
- Understand the importance, meaning and writing of each paper section.
- Know some hints on scientific writing.

1 Scientific texts

There is a large variety of scientific texts: from reporting the results achieved during a long term research project, to briefly discussing on specific results published in a journal. Although different types of scientific texts do have specific length or purpose, they all have a quite similar structure. For instance, almost all scientific texts include a bibliography section, an introduction or are summarised by means of the abstract. Moreover, authors make use of some language conventions when writing scientific texts. Brevity and clarity are the basis of scientific writing style: the goal in writing is to achieve objectivity, which is far from using language artifacts and pomposity.

In this section, we briefly introduce the scientific paper and other kinds of scientific texts.

1.1 The scientific paper

Research results have to be published as soon as possible. It will clearly allow other researchers to know the newest advances on a topic. In that sense, scientific results are published in form of scientific paper.

In general, the goal of a scientific text is to describe processes and results which take place when doing research: the results of a new method to control cancer growing, the results in terms of processing time of an improvement in a greedy algorithm, etc.

> A scientific paper or article is a scientific text with a well established structure, whose goal is to describe a research advance.

As in most of scientific texts, papers must justify the reason for the described research (i.e. why need their authors present a new proposal) and must demonstrate the soundness of the research (in terms of experimental results or analytical proofs).

Scientific papers follow a conventional structure:

- An abstract to briefly summarise the whole paper.
- An introduction to the problem to be tackled (pointing out what will be done and what will be achieved)
- A background including a review of existing proposals.
- A section explaining the new proposal.
- A section assessing the method and/or comparing the new proposal with existing ones. All the goals said to be achieved must be assessed.
- A section concluding the paper (summarizing what has been done and what has been achieved).

Conferences and scientific journals.

Authors may present their works in a conference or may try to publish them in a scientific journal. During the publishing of a paper, the editor of the journal will make a final decision on the paper to publish; in a conference, the programme committee chair makes the same function. Finally, the publisher will typeset the accepted papers and print the journal. Regarding conferences, the papers are usually published in a book of proceedings. We will elaborate on these topics in Module X.

- A bibliography section.

There may be some other parts in a paper (for example, an acknowledgements section). Later in this module we will describe the structure and the writing of a paper.

Figures 1 and 2 show a 4-pages example paper with its different sections. The example paper is structured as follows:

- The mandatory title, authors and abstract.
- A section introducing the topic and technology the paper elaborates on (Section 1).
- A section with background information on previous work in location privacy (Section 2).
- A section where the new proposal is described (Section 3).
- A section discussing on the validity on the new proposal is presented (Section 4).
- A section concluding and pointing out new research lines (Section 5).
- Some acknowledgements and the bibliography section (Acknowledgements section and References section, both unnumbered).

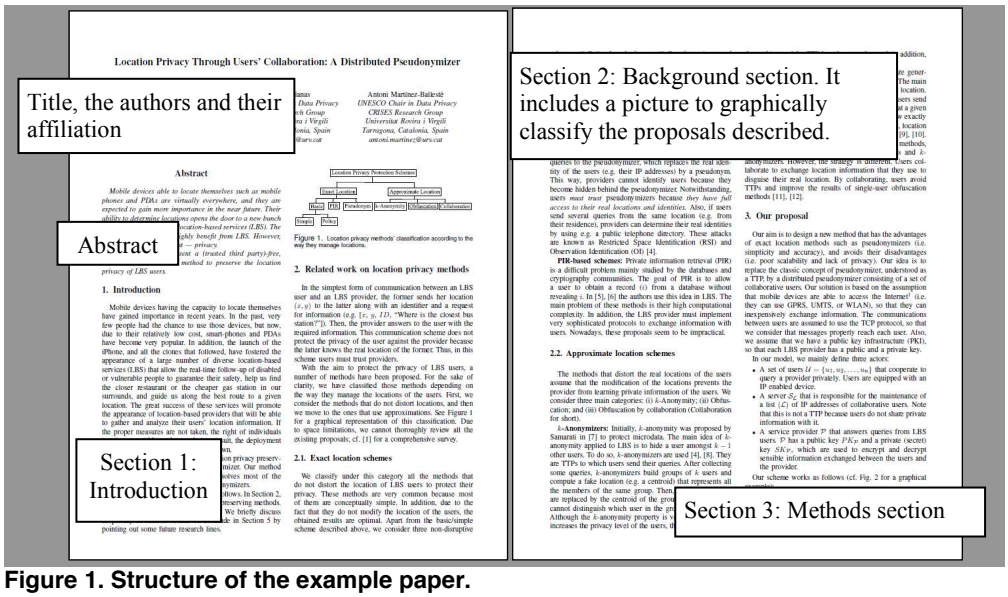


Figure 1. Structure of the example paper.

The reference of the example paper

P. A. Pérez-Martínez, A. Solanas and A. Martínez-Ballesté "Location Privacy Through Users' Collaboration: A Distributed Pseudonymizer". Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies - UBICOMM 2009, pages 338-341. Sliema, Malta. Oct 2009. It can be obtained from http://dx.doi.org/10.1109/UBICOMM.2009.60

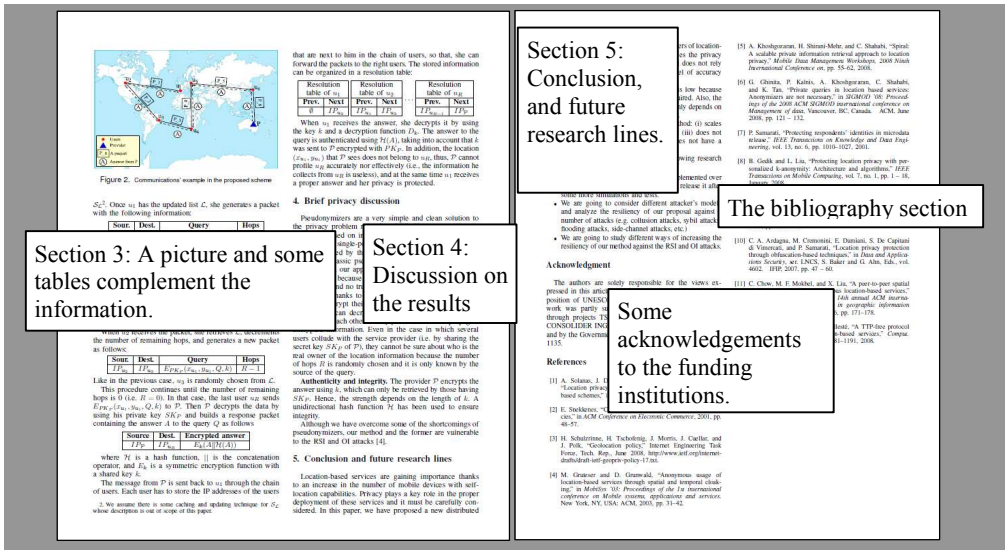


Figure 2. Structure of the example paper.

The first thing an author should have in mind when writing a paper is the motivation of the paper itself. Hence, authors should ask the next questions before starting to write, since they implicitly describe the meaning of the scientific paper:

- Why do we write the paper? Have we come up with a new idea? Have we improved any existing proposal?
- How shall we explain our work? Can we prove we are right?
- Has anybody published the same idea before? If so, can our approach improve the published one?

In the next lines, we discuss on the previous questions. If you have a new idea (for instance, using data perturbation for location anonymity) you should search for similar existing proposals. In that sense, you should take a look at the scientific literature to know if anybody has already published the same idea. You can use search engines (or even scientific literature search engines) to look for similar proposals. You could start writing something like "location anonymity", and you will presumably receive a list of scientific contributions dealing with location anonymity. You have to collect any reference concerning the same topic of your proposal: when writing a paper you have to show that, as far as you know, no one has come up with the same idea before.

Google Scholar

The Google's search engine for academic publishing can be a good starting point to find out if someone already had the same idea than you.

If no one has come up with the same idea, you can write your method and do some experiments or formal proofs to assess its validity. If there are some existing proposals on your topic, read them carefully: maybe your approach is better than all the existing ones!

Describing your approach and trying to assess its validity is a tricky part. On the one hand, you may not have plenty of time to implement a prototype for your idea or processing thousands of output data. On the other, your proposal can present some flaws

which cannot be seen at a first glance. Hence, you can look for a co-author among the personnel in your research group. If you discuss your proposal with a colleague, it will be free of flaws or even will improve considerably. Moreover, if this colleague helps you with the implementation or the analysis of results, the work will be done in a half of the time. Note that there may be someone else writing the same idea on a scientific paper. Consequently, time is crucial.

During the rest of this module, we elaborate on writing scientific papers. However, we must pay attention to other kind of scientific texts.

1.2 Other scientific texts

There are some other kinds of scientific texts. They all have a quite similar structure. However, according to the length, there are long scientific texts (such as books or thesis, that will be divided into chapters and/or parts), and short scientific texts (such as papers).

Information Processing Letters

To illustrate what is a letter, we refer to the aim of the Information Processing Letters Journal "The aim of Information Processing Letters is to allow rapid dissemination of interesting results in the field of information processing in the form of short, concise papers."

- **Report.** It is usually longer than a research paper and may contain preliminary work. Hence, a report may be used for reference in a research group or community of researchers, but not considered for publication in form of scientific paper. However, if the content of the report is to be published as paper, some parts may be suppressed in order to accommodate a specific length. Research projects involve the writing of long research reports in order to describe the results achieved within the research project. If reports are long, they are usually divided into chapters.

- **Survey paper.** It reviews and compares the work of other scientists in order to come up with future trends in their research fields. Hence, they do not include real new proposals, but objectively compare previous ones. When there are several proposals on a topic, it may be interesting to compare them. The contribution sections of a survey paper deal with the deep analysis of the literature surveyed. The results section must provide the reader with an accurate comparison between the studied proposals.

- **Position paper.** These papers are shorter than survey papers. The existing proposals are just briefly reviewed (without a methodic comparison between them) and authors point out the research lines on the topic that scientists should follow.

- **Letter.** It is a very short paper, usually written to communicate a proposal that does not need the extension of a paper. Moreover, letters may review other papers which are already published.

- **Scientific book.** It consists of a set of chapters elaborating on a specific research field. A researcher expert in a field asks for contributions to other colleagues which are researching on the same field. The researcher acts as editor of the book and selects which contributions will appear in it. The final

result is a compilation of the recent advances in the research field, and can be of utility to researchers willing to have a general picture of the topic the book focuses on. Another typical book in scientific writing is the **book of proceedings** of a conference: it is composed by all the papers presented in a conference. If only the abstracts of the contributions are to be published rather than the whole papers, the book is then a **book of abstracts.**

- **Ph.D. thesis.** A thesis or dissertation is a usually large and deeply-elaborated text which is mandatory for obtaining a M.Sc. or Ph.D. degree. It synthesises the work that the candidate for the degree has been developing. A M.Sc. thesis elaborates on a very specific topic and will presumably involve the work of several months. On the contrary, the thesis for obtaining the Ph.D. is the result of a long term research (usually several years). The structures of both theses are similar to that of scientific papers. It includes an introduction, a review of the relevant literature on the field, a set of chapters describing the research done (since the Ph.D. involves more research than an M.Sc., the author is expected to write several chapters), a summary of results, conclusions and a bibliography.

M.Sc. and Ph.D.

M.Sc. stands for Master of Science and is the first academic degree a postgraduate can earn.

Ph.D. stands for Philosophiae Doctor and is the highest academic degree one can earn.

The hints and techniques explained during the rest of the module are not only valid for scientific papers but also for any kind of scientific text.

2 Writing a scientific paper

In this section we describe the parts of a scientific paper. We elaborate on their writing, and give several examples.

The structure of a scientific paper, which includes the parts introduced in section 1.1, follows the IMRAD convention, where:

- I stands for Introduction (Section 1 and 2 of the example paper).
- M stands for Methods (Sections 3 of the example paper).
- R stands for Results (Sections 4 and 5 of the example paper).
- And D, that stands for Discussion (Sections 4 and 5 of the example paper).

The final number of sections and their titles are not mandatory (except for some specific publishers) and hence the IMRAD structure does not exactly fit the structure of the example paper. However, the basic idea of the IMRAD structure is that a paper must be structured as any other story: opening, development and conclusion.

In this section, we elaborate on describing the parts of the scientific paper.

2.1 Title

The literature

The word literature is often used for generally describing the set of all scientific papers.

The title identifies the paper. Hence, it must be complete and original enough to avoid new papers having the same title as existing literature. Moreover, your title should provide as much relevant information as possible. For instance

Privacy Preserving Techniques

is likely to be a bad title for a paper, since it does not give many details. With this title, one will expect a paper about all the methods for privacy preservation. On the contrary, the title

Privacy Preserving Techniques in Statistical Databases

gives the reader more information. However, do authors present a new technique for privacy preservation in statistical databases? Or is it just a survey of existing methods? Next, we show the two options:

Comparison of Privacy Preserving Techniques in Statistical Databases

A New Method for Privacy Preservation in Statistical Databases

The latter is likely to deal with a new algorithm for privacy preservation but, what happens if there already exist tenths of different proposals that achieve the same privacy-preserving goal? Authors should specify in the title why their approach is important. In other words, the title must summarise your proposal. For instance, the title

A New Method for Privacy Preservation in Statistical Databases Based on Improving Microaggregation

tells the readers that the proposal improves microaggregation. However, if authors present their work in a conference about privacy in statistical databases whose attendants are really experts and are aware of all the microaggregation techniques, the next title seems more suitable for the paper:

An MDAV Based Approach for Near-Optimal Microaggregation in Numerical Databases

MDAV

Certainly, MDAV is a microaggregation technique: Maximum Distance to Average Vector. Hence, is there any problem with using abbreviations in the title? If the abbreviations are well-known to your audience, there should be no problem.

Now, we can examine the title of the example paper:

Location Privacy Through Users' Collaboration: A Distributed Pseudonymizer

It gives us the following information:

- It is a novel approach whose goal is location privacy.
- It is achieved by means of the collaboration of the users.
- The protocol presented acts as a pseudonymiser and it is a distributed architecture.

Finally, regarding the example title, it is often interesting to use compound titles. A simple title would perhaps result less "commercial":

A Distributed Anonymizer for Location Privacy Through Users' Collaboration

2.2 Authors

You have an idea and you write a paper, hence you are an author. However, if you ask other colleagues for help, which is the frontier that defines who has earned appearing as coauthor? First, we can think of a list of tasks involved in writing a paper:

- **Having the main idea.** This is important, since without idea there is no paper.
- **Doing some experiments**. In disciplines such computer science, this will involve some programming.

- **Doing an exhaustive search of the literature** in order to find proposals similar to yours and summarise them.

- **Writing the paper.** This is a difficult skill, especially for those researchers not having English as their mother tongue. Writing the paper also involves drawing pictures, typesetting some tables, etc.

- **Reviewing the paper.** It is usual that you request a reading to your colleagues: this is useful not only for finding writing mistakes, but also to assess that our idea is perfectly transmitted with the reading of the paper.

Certainly, all these tasks can be carried out by a single researcher. However, research tasks are usually a group work (at least, a work between Ph.D. students and their advisors!) Moreover, when you explain your idea to other members in the research group, some colleagues may come up with fresh ideas and improvements.

As a result, the final number of authors depends on several facts. On the one hand, in some disciplines, all the researchers working under a research project collaborate in doing an experiment and, thus, they all generally become authors of the papers. However, it is possible that the importance that a paper has in your curriculum depends on the number of authors: in that sense, you may not be interested in sharing authorship with all the colleagues of your research group! Hence, deciding authorship is always tricky. A usual tradeoff consists of becoming author if and only if you have worked hard in the paper (that is, have got the idea and written it down or have developed the experiments) and the colleagues who may have reviewed the paper may appear in an acknowledgements section.

Acknowledgements

We elaborate on this section below.

If deciding who becomes an author is not straightforward, deciding the order of the authors is not either an easy task. Hence, the alphabetical order is commonly used when all authors have contributed in a similar way.

The name of each author should be accompanied by his or her affiliation (that is, the institution the author belongs to, its address, e-mail address and other details). Moreover, sometimes just the address of the main author is specified. In this case, it is usual that the authors' line only contains their names whereas a footnote in the name of the main author contains the address of this so-called corresponding author. However, depending on the typesetting format used by the publisher, the affiliations of all authors will be placed as footnotes.

Corresponding author

The corresponding author is usually the main author of the paper, who is also responsible for following all the processes that the paper will follow after being finished. We will elaborate on these processes in the module Publishing the Research.

Last but not least, the name of the author is another important issue. Search engines of scientific databases allow the search using the author name. Hence, the first time someone becomes the author of a paper, the name itself must be clearly decided. For instance, imagine how many authors have a name similar to John Smith or José Garcia. Concerning Spanish author names, it is usual that you make use of your second surname, e.g. Antoni Martínez-Ballesté, Jordi Castellà-Roca, Josep Domingo-Ferrer, etc. Note that

a dash is used to maintain both surnames tight: if a blank was used as separation between surnames, English search engines would index these authors as Antoni M. Ballesté, Jordi C. Roca and Josep D. Ferrer. Another aspect to take into account is to use accents or not. Some authors prefer not to use accents in order to simplify database indexation and search. However, in case you have an uncommon first surname, you may decide not to use your second surname.

2.3 Abstract

The first text that appears below the title and authors is the abstract. It consists of a short introduction summarising the article (hence, it is sometimes called summary). Its length is usually between 50 and 300 words (these limits may be specified by the publisher).

> The abstract is a summary of the whole paper. Hence, it must briefly introduce the topic and point out the main achievements.

The abstract must contain:

- The main idea the paper elaborates on.
- A brief description of what is done.
- A brief description of the results (if any).

The abstract is usually the only part of the paper which is freely accessible via an Internet research repository: only subscribers may have access to the full content of the paper (see Figure 5). Hence, the abstract should make decide a potential reader between paying a fee for the full paper or not. Moreover, the abstract may be the only part of the paper that busy readers will read: so it must convince them to read the full paper. Since the abstract can be published on its own, it cannot contain any references (to figures in the paper or to bibliographic cites). Finally, the abstract is written in present tense.

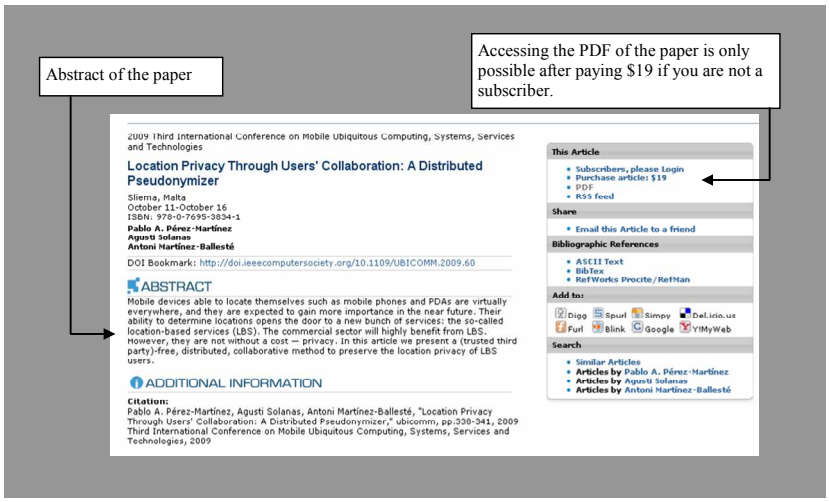


Figure 3. Accessing the example paper via the IEEE Computer Society digital library service.

In the example paper, the abstract introduces the problem tackled and briefly describes the proposed solution.

> Mobile devices able to locate themselves such as mobile phones and PDAs are virtually everywhere, and they are expected to gain more importance in the near future. Their ability to determine locations opens the door to a new bunch of services: the so-called location-based services (LBS). The commercial sector will highly benefit from LBS. However, they are not without a cost — privacy. **(the problem)**
>
> In this article we present a (trusted third party)- free, distributed, collaborative method to preserve the location privacy of LBS users. **(the solution)**

In the next example, the abstract certainly (1) summarises the aim of the work, (2) describes the methodology and (3) summarises the results:

> The fast growth of Radio Frequency IDentification (RFID) implies a deployment challenge, that is how to keep this technology scalable without renouncing security and privacy features. This paper focuses on combining tag privacy and scalability using the hash locks mechanism. **(1)**
>
> Our contribution is twofold: (i) a cell-based architecture is proposed in which the readers co-operate in order to conduct tag identification in a private and scalable way; (ii) a communication protocol engineering the proposed architecture is defined and assessed.**(2)**
>
> The proposed architecture and protocol solve the scalability shortcomings of private RFID identification pointed out by Juels and Weis.**(3)**

After the abstract, the publisher may ask for a list of key words that allow the classification of the paper. These are chosen by the authors among those proposed by the publisher of the paper.

2.4 Introduction

Once the title, authors, abstract and keywords have been specified, the body of the paper starts with the first section, simply called Introduction.

> The main goal of the Introduction section is to focus on the topic the paper elaborates on.

The Introduction section has three objectives:

- First, it must to introduce the topic (i.e. Which is the theme of the paper?) This introduction to the topic must be more or less exhaustive depending on the audience of the paper. Clearly, if your paper is to be published in a Computer Engineering journal and your paper is about the use of a database, you should not talk about what a database is!

The introduction of a survey paper

If you are writing a survey paper, the introduction will just elaborate on the topic background. Instead of solving problems, the aim of the paper will be to do a comparison/classification of the existing proposals for the topic.

Motivation

The problems and their solutions you write in the introduction are, in fact, the motivation of the paper. Hence, the Introduction section is sometimes entitled Motivation.

- Second, it must explain which problems you have found or which problems are well-known for the topic.

- Finally, you have to point out the main goals of the paper (which will presumably include tackling the aforementioned problems).

This is the introduction section of the example paper. First it introduces the technology (LBSs) and then points out their problem related to privacy of users:

> Mobile devices having the capacity to locate themselves have gained importance in recent years. In the past, very few people had the chance to use those devices, but now, due to their relatively low cost, smart-phones and PDAs have become very popular. In addition, the launch of the iPhone, and all the clones that followed, have fostered the appearance of a large number of diverse location-based services (LBS) that allow the real-time follow-up of disabled or vulnerable people to guarantee their safety, help us find the closer restaurant or the cheaper gas station in our surrounds, and guide us along the best route to a given location. The great success of these services will promote the appearance of location-based providers that will be able to gather and analyze their users' location information. If the proper measures are not taken, the right of individuals to privacy could be endangered. As a result, the deployment of LBS could be remarkably slowed down.

After the topic is introduced, the goals of the paper are summarised:

> In this article, we propose a new location privacy preserving method --- A distributed pseudonymizer. Our method is based on users' collaboration and solves most of the drawbacks of classic centralized pseudonymizers.

And the organisation of the paper is described:

> The rest of the article is organized as follows. In Section 2, we summarize several location privacy preserving methods. We describe our proposal in Section 3. We briefly discuss its privacy in Section 4, and we conclude in Section 5 by pointing out some future research lines.

If the introduction is long and has been divided into subsections, the two last points are detailed in a specific subsection called “Contribution and Plan of this Paper” in which authors detail the goals of the paper and also describe its structure.

2.5 Background and related work

Once the topic has been introduced, a section with background information must be written in order to:

- Describe the background concepts that are essential to understand your achievements.

- Highlight your proposal among the existing ones.

This section may be simply entitled "Background". However, if the background includes information on several previous proposals or includes detailed up to date information about a certain topic or technology, the section can be entitled "Previous Work", "Related Work" or "State of the Art".

While depicting previous proposals, authors must have in mind which aspects (or shortcomings) of these existing works are improved by the new proposal. Concentrating on these aspects will make your proposal more interesting.

During this section, several papers by other authors will be cited (naturally, some previous works by the authors of the paper may be also cited). Thus, one question that comes to mind is what to explain about the reviewed previous works. A common practice is to give just a succinct description of the proposal, its bibliographic citation, and briefly point out its main features.

Citations

We elaborate on bibliographic citations below on this module.

If the list of previous proposals is too long, it is a common practice to summarise them using some kind of classification. For instance, see this excerpt from the Related Work section of the example paper:

> ***k*-Anonymizers:** Initially, k-anonymity was proposed by Samarati in [7] to protect microdata. The main idea of k-anonymity applied to LBS is to hide a user amongst k-1 other users. To do so, k-anonymizers are used [4,8]. They are TTPs to which users send their queries. After collecting some queries, k-anonymizers build groups of k users and compute a fake location (e.g. a centroid) that represents all the members of the same group. Then, the real locations are replaced by the centroid of the group and the provider cannot distinguish which user in the group sent the query. Although the k-anonymity property is very interesting and increases the privacy level of the users, this approach has all the problems of the TTP-based approaches and, in addition, the obtained results are not accurate.

The text above:

- Introduces a group of location privacy previous proposals: k-anonymisers.
- Briefly describes the properties of k-anonymisers.
- Briefly points out their shortcomings.

When describing related work, it is sometimes useful to summarise the concepts in a figure or using a table. For instance, in the example paper the scheme of Figure 4 is used.

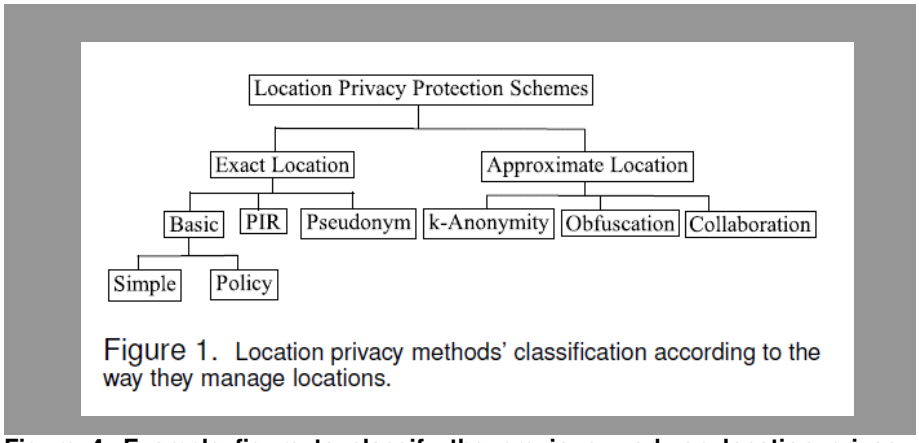


Figure 4. Example figure to classify the previous work on location privacy methods.

As a conclusion, the length and the depth of the background section will certainly depend on several aspects, e.g. the quantity of existing proposals, the assumed knowledge on the topic of the audience of the paper, etc. If the editor requests authors to shorten their paper for some reasons, the background section may be the best candidate. Moreover, if the amount of information to be written in the background section is not enough for a whole section, the background information is placed in the Introduction.

2.6 Describing your work

The description of your proposal deserves a whole section. In some scientific disciplines, this section is usually called "Materials and Methods". Regarding Information Technologies disciplines, it is not common to use this title. Perhaps, this would be appropriate in case of using a set of hardware with several requirements and specifying an algorithm to execute. The section is hence called "Our Proposal", "Our Approach", etc. or simply using a description of what you are going to present (for example "A New Method for RFID Secure Identification").

When describing your proposal there are many ingredients to use and several ways to mix them. The quality of the section and consequently the success of the authors explaining their proposal to the scientific community depend on the ability of authors. In Information Technology related disciplines, these are basically the kinds of elements a paper can present:

- An **algorithm**, that presents a new method or a variation of an existing one.
- A **protocol**, which consists of an algorithm describing the communication and behavior between two or several parties.
- An **architecture**, which depicts a more or less complex system of components. Its description may certainly involve the definition of algorithms and protocols between its components.

- Some theoretical results, which are given in form of **theorems**, **lemmas** and **corollaries**.

In general, several of the above elements are likely to be used in the paper. When describing the main topic or topics of the paper, it is essential to use **definitions** to clarify all the elements that take part in the proposal. In addition, the description of a proposal is usually complemented by the use of pictures: e.g. the use of **schemes** to represent circuits, the use of **flowcharts**, etc.

In our example paper, some definitions are introduced prior to describing the proposal:

> In our model, we mainly define three actors:
>
> - A set of users $U=\{u_1, u_2,\ldots,u_n\}$ that cooperate to query a provider privately. Users are equipped with an IP enabled device.
> - A server S_L that is responsible for the maintenance of a list L of IP addresses of collaborative users. Note that this is not a TTP because users do not share private information with it.
> - …

In the following paragraphs some of the techniques for describing the above elements are presented. Some toy examples are to be used.

Describing an algorithm

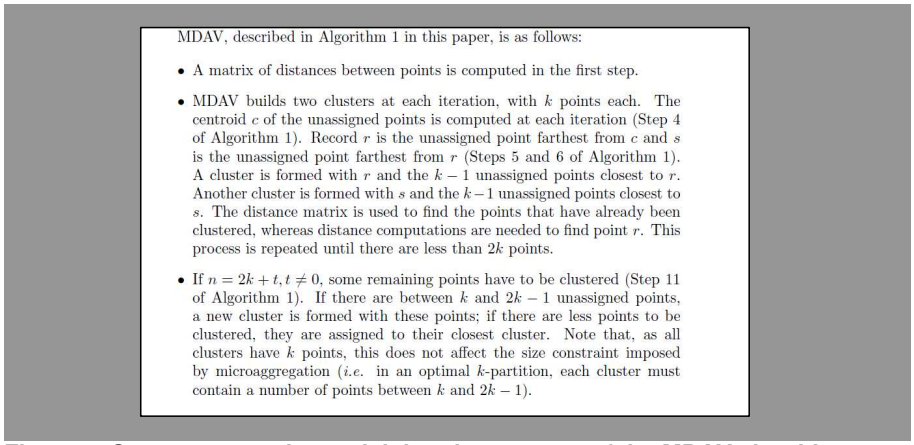
Before describing the algorithm, some of the elements that take part in it must be defined (see the set of users *U* or the server defined in the above example). Moreover, in some cases the definition can be inserted in the algorithm description itself.

There are basically two ways of describing an algorithm:

- Elaborating on a list of steps that constitute the algorithm. In this case, the algorithm is just described. The sentences corresponding to the steps are usually numbered. However, some authors may prefer to merely describe the algorithm in a single paragraph, although it seems a less elegant manner of describing algorithms.

- Describing the algorithm using pseudocode rather than a specific programming language.

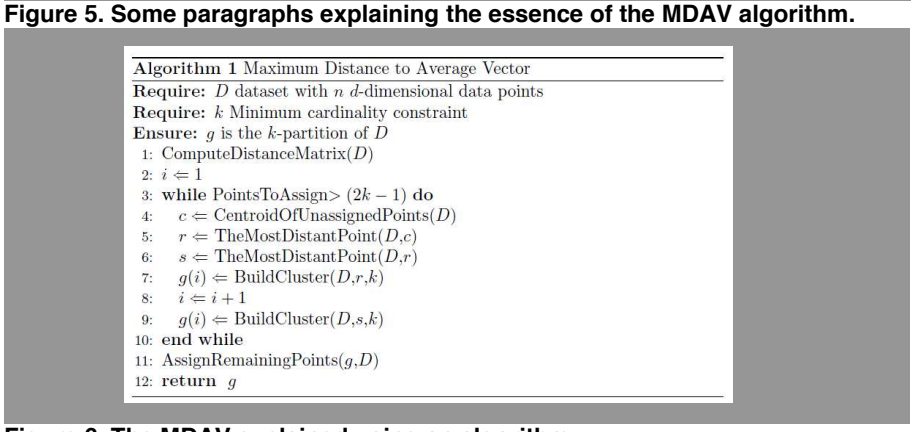
Both approaches are valid and widely used: the key point is to be precise and concise. Moreover, both techniques can be used in the same paper: one can use a text describing the algorithm while referring to its "algorithmised" version. To illustrate both ways of describing an algorithm, Figure 5 shows the explanation of the MDAV algorithm using a set of sentences. Figure 6 shows an algorithmic explanation of MDAV.



MDAV, described in Algorithm 1 in this paper, is as follows:

- A matrix of distances between points is computed in the first step.
- MDAV builds two clusters at each iteration, with k points each. The centroid c of the unassigned points is computed at each iteration (Step 4 of Algorithm 1). Record r is the unassigned point farthest from c and s is the unassigned point farthest from r (Steps 5 and 6 of Algorithm 1). A cluster is formed with r and the $k-1$ unassigned points closest to r. Another cluster is formed with s and the $k-1$ unassigned points closest to s. The distance matrix is used to find the points that have already been clustered, whereas distance computations are needed to find point r. This process is repeated until there are less than $2k$ points.
- If $n = 2k + t, t \neq 0$, some remaining points have to be clustered (Step 11 of Algorithm 1). If there are between k and $2k-1$ unassigned points, a new cluster is formed with these points; if there are less points to be clustered, they are assigned to their closest cluster. Note that, as all clusters have k points, this does not affect the size constraint imposed by microaggregation (*i.e.* in an optimal k-partition, each cluster must contain a number of points between k and $2k-1$).

Figure 5. Some paragraphs explaining the essence of the MDAV algorithm.



Algorithm 1 Maximum Distance to Average Vector

Require: D dataset with n d-dimensional data points
Require: k Minimum cardinality constraint
Ensure: g is the k-partition of D

```
ComputeDistanceMatrix(D)
i ⇐ 1
while PointsToAssign> (2k − 1) do
    c ⇐ CentroidOfUnassignedPoints(D)
    r ⇐ TheMostDistantPoint(D,c)
    s ⇐ TheMostDistantPoint(D,r)
    g(i) ⇐ BuildCluster(D,r,k)
    i ⇐ i + 1
    g(i) ⇐ BuildCluster(D,s,k)
end while
AssignRemainingPoints(g,D)
return g
```

Figure 6. The MDAV explained using an algorithm.

Describing a protocol

A protocol can be seen as an algorithm that implies the communication between two parties. Hence, describing a protocol must concentrate on the messages. In that sense, authors must specify:

- The content (i.e. the fields) of the messages.
- The flow of the messages.

A plethora of techniques can be used to depict the messages and how they are sent and received during the protocol. A message can be described using a figure, a table or just using an expression specifying the content of the message. On the other hand, the flow of the messages can be illustrated by means of a figure. However, note that the text in the article must refer to the figure and elaborate on the description of the protocol. Figure 7 shows the description of a simple example protocol between two parties. The messages have been described using expressions and the message flow is described using a picture.

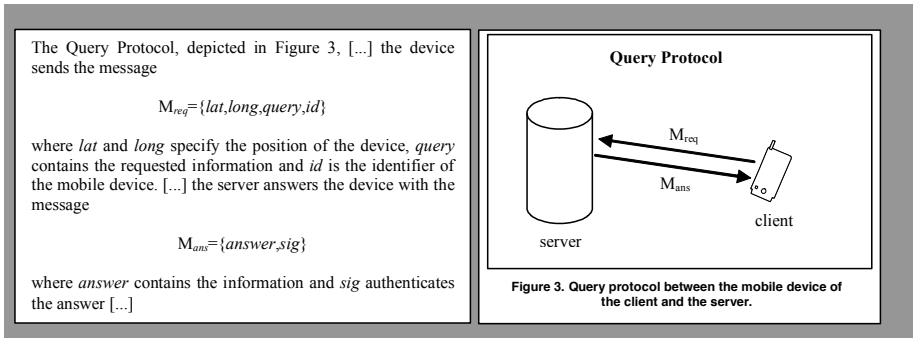


Figure 7. Example of a protocol described by means of text and a picture.

Describing an architecture

By architecture we refer to a complex system that involves several parties. In Figure 8, an example architecture is depicted. One would expect that the paper defines each block of the system and describes its behaviour (using algorithms and protocols if necessary). However, a picture illustrating the whole system is usually mandatory when your paper elaborates on an architecture.

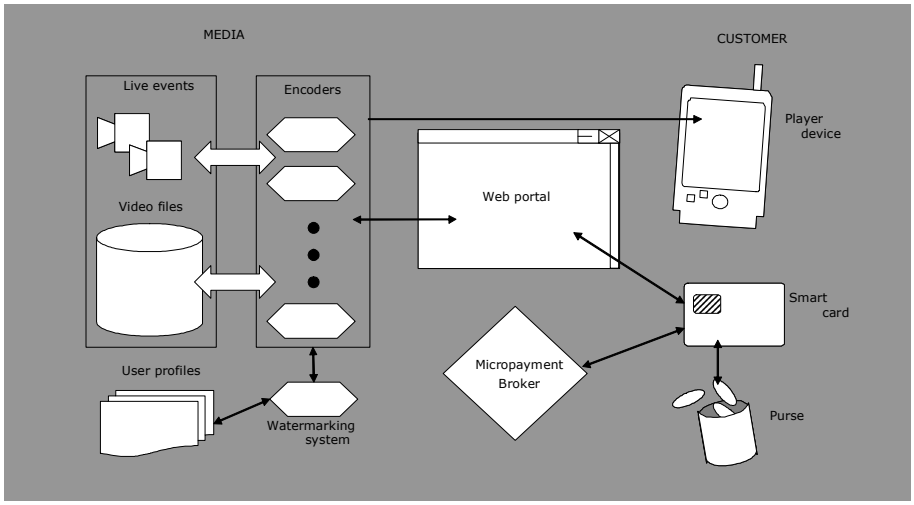


Figure 8. Example of simple architecture.

Formulae, theorems and other resources

Theoretical resources are commonly used in scientific disciplines related to mathematics. However, formulae, analytical expressions, lemmas and other tools are not prohibited in technology-related papers and always paint them with some elegance.

In fact, there are plenty of chances of using expressions so as to express actions and define terms. In one of the above examples, during the description of the MDAV algorithms, the authors have used the following expression:

$$\text{If } n=2k+t,\ t\neq 0 \text{ some remaining points have to be clustered…}$$

Moreover, in Figure 7 there was a set of users defined using an expression:

$$U=\{u_1,\ u_2, \ldots, u_n\}$$

In the latter example, the expression is placed in a single line because the authors want to remark it.

Last but not least, we must mention another elegant set of tools whose use is mandatory in some disciplines such as mathematics or cryptography. Definitions, lemmas, claims, theorems are often used for proving theoretical statements:

- As stated above, the actors and elements which take part in algorithms or protocols must be defined. Hence, the elements needed in the theoretical statements are also defined by means of **definitions**.

- If the theoretical statement is not straightforward to prove, it is considered a **theorem**. This term is reserved for important theoretical statements. Hence, to prove a theorem it must be split into smaller theoretical statements which have to be proved individually.

- These smaller theoretical statements can be either **lemmas** or **claims**, the latter being a kind of "unformal" lemmas. Thus, once the set of lemmas leading to the theorem have been proved, authors are then in condition for defining and proving the theorem. Figure 9 depicts a scheme of proving a theorem by means of lemmas.

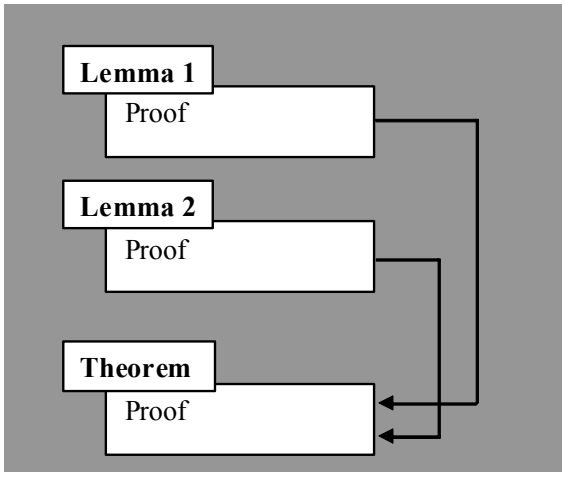


Figure 9. A theorem being proven by means of the previous statement of two lemmas.

2.7 Presenting results

After the main proposal of the paper is presented, authors should demonstrate the correctness of their research somehow. In topics related to mathematics, it is common to demonstrate concepts by means of lemmas and theorems. Some specific proposals, such as simple protocols or algorithms, do not need to be demonstrated. In these cases, the achievements can be demonstrated simply elaborating on how these achievements are met.

However, in most scientific and technical disciplines, the validity of the proposals is shown by means of experiments, simulations and tests.

Presenting results is a matter of being brief and clear. Hence, it is not a good practice to provide readers with much overwhelming information. Moreover, the presentation of the data, generally using tables and charts, must be designed accurately enough so as to reflect the main results. Finally, the captions used to describe the table or the chart must provide readers with enough information to understand the data.

Regarding tables:

- They should be arranged so as to highlight the most significant results (for example, using bold typeface).
- They should not include redundant or unnecessary data.
- They should specify the units and/or statistical notation for the values.

Regarding charts:

- Authors should accurately choose the type of chart.
- Numbers should not be included when they are not important (the axes of the chart approximately inform of the values and differences between bars allow comparisons).
- The axes should be accurately designed.
- The title and legend should provide readers with enough information.
- If the printing is to be in monochrome different type of lines, symbols and patterns must be used instead of colors.

To illustrate some of the above items, Figures 10 and 11 show a chart with some problems and an improved version respectively.

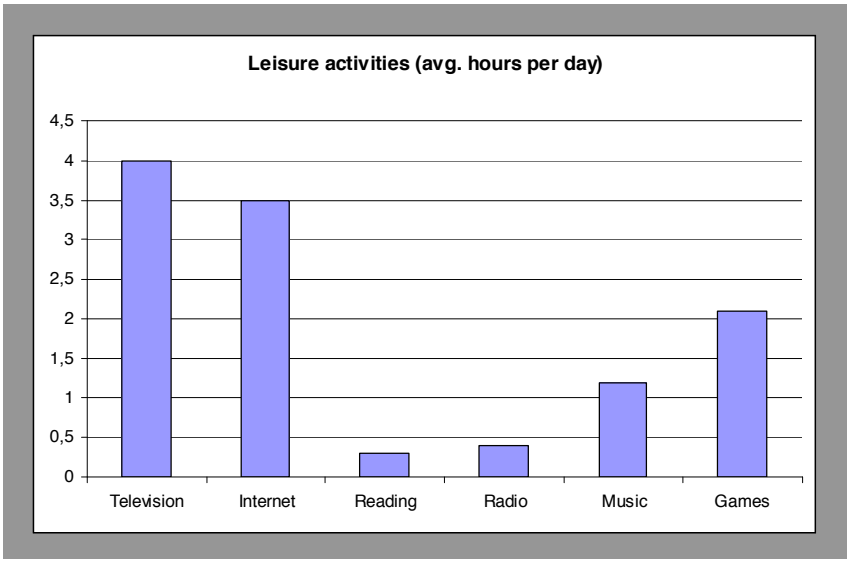


Figure 10. This chart has a problem with the vertical axis. Moreover, columns could be ranked.

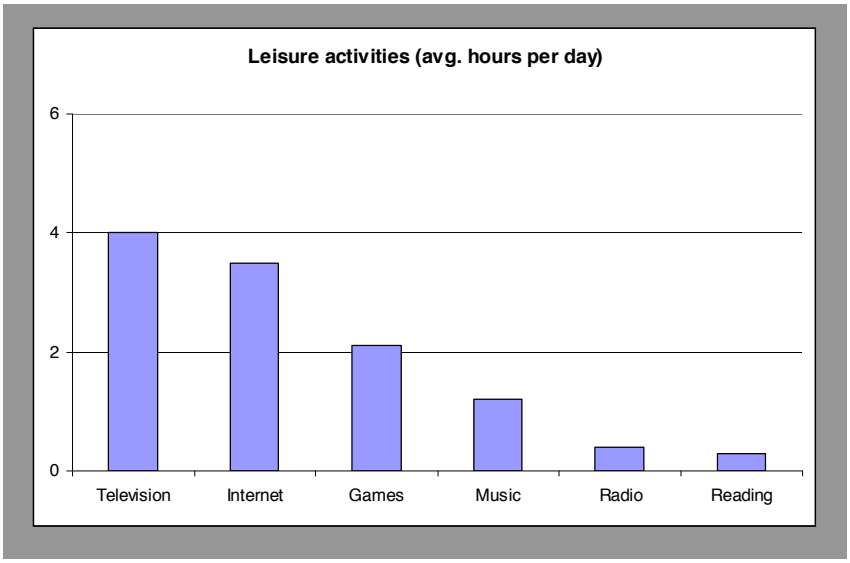


Figure 11. An improved version of Figure 10.

2.8 Discussion and conclusions

Concluding a paper has the aim of discussing on the assessment of the goals the authors pointed out in the introduction section. It must summarise what we the authors presented, and hence it must be written in past tense or in present perfect (remember that the abstract is written in present). Since authors assume the reader has read the above parts of the paper, they can briefly elaborate on the way goals have been achieved. Moreover, since the conclusion can contain the discussion on if the goals of the paper have been achieved, the section can be entitled "Discussion". In the discussion section, authors may also compare the proposal of the paper with those addressed in the background section.

Since the results typically open the door to further research on a topic, some hints on the future research are usually pointed out. Hence, the concluding section may be entitled "Conclusion and Future Work" or "Conclusion and Further Work". It is a good place to recall all these aspects that, due to space constraints, have not been tackled in the presented paper.

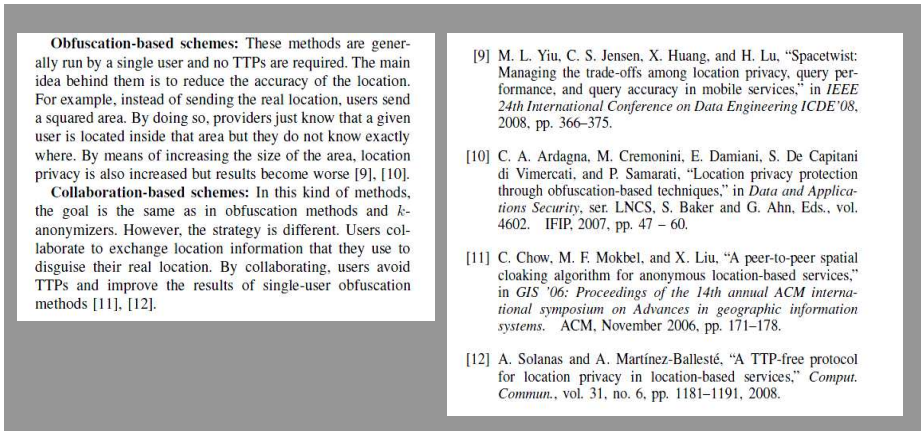
There is no mandatory structure for the "ending" of a paper. However, the body of the paper must end with a discussion on the results and with some approaches to the future work.

As an example, this is the concluding section of the example paper. The first three paragraphs elaborate on summarising the achievements of the paper. The last part is devoted to sketching the future research lines.

> Location-based services are gaining importance thanks to an increase in the number of mobile devices with self-location capabilities. Privacy plays a key role in the proper deployment of these services and it must be carefully considered. In this paper, we have proposed a new distributed pseudonymizer to protect the privacy of the users of location-based services. Our proposal clearly improves the privacy level achieved by classic pseudonymizers (it does not rely on a TTP) whilst maintaining the same level of accuracy (the location is not distorted).
>
> The complexity of the proposed method is low because only simple cryptographic operations are required. Also, the communications overhead is quite low and only depends on the number of users involved in the protocol.
>
> Thanks to the avoidance of TTPs, our method: (i) scales properly; (ii) does not generate bottlenecks; (iii) does not require users to trust anybody; and (iv) does not have a single point of attack.
>
> In the future we plan to address the following research lines:
>
> - We have a prototype of our proposal implemented over the Iphone platform and we are going to release it after some more simulations and tests.
> - We are going to consider different attacker's models and analyze the resiliency of our proposal against a number of attacks (e.g. collusion attacks, sybil attacks, flooding attacks, side-channel attacks, etc.)
> - We are going to study different ways of increasing the resiliency of our method against the RSI and OI attacks.

2.9 Bibliography

In the body of the paper, the authors cite other papers, textbooks, websites, etc. These citations must be referenced, i.e. the author, the name of the paper, the way to find it, etc. must be specified in a Bibliography or References section. In some of the used examples, some cites can be found. Figure 12 presents an excerpt of the example paper with some citations and their corresponding bibliographic references.



Obfuscation-based schemes: These methods are generally run by a single user and no TTPs are required. The main idea behind them is to reduce the accuracy of the location. For example, instead of sending the real location, users send a squared area. By doing so, providers just know that a given user is located inside that area but they do not know exactly where. By means of increasing the size of the area, location privacy is also increased but results become worse [9], [10].

Collaboration-based schemes: In this kind of methods, the goal is the same as in obfuscation methods and *k*-anonymizers. However, the strategy is different. Users collaborate to exchange location information that they use to disguise their real location. By collaborating, users avoid TTPs and improve the results of single-user obfuscation methods [11], [12].

[9] M. L. Yiu, C. S. Jensen, X. Huang, and H. Lu, "Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services," in *IEEE 24th International Conference on Data Engineering ICDE'08*, 2008, pp. 366–375.

[10] C. A. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and P. Samarati, "Location privacy protection through obfuscation-based techniques," in *Data and Applications Security*, ser. LNCS, S. Baker and G. Ahn, Eds., vol. 4602. IFIP, 2007, pp. 47 – 60.

[11] C. Chow, M. F. Mokbel, and X. Liu, "A peer-to-peer spatial cloaking algorithm for anonymous location-based services," in *GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*. ACM, November 2006, pp. 171–178.

[12] A. Solanas and A. Martínez-Ballesté, "A TTP-free protocol for location privacy in location-based services," *Comput. Commun.*, vol. 31, no. 6, pp. 1181–1191, 2008.

Figure 12. Example of citations and their bibliographic references.

There are several ways of citing a work inside the text. The most common in Information Technologies disciplines is using numbers. In **numerical** citations a number refers to the bibliographic item in the bibliography section. Some examples follow: for a single cite [5], for several cites [5, 7, 19], for contiguous cites [1-5] (meaning from 1 to 5). We can also use a mixture of them [1-5,9]. Of course, in the bibliography section, references will be numbered accordingly. However, when using numerical cites and regarding their sorting:

- Bibliographic references can be sorted according to the order they are cited in the paper (hence the first cite will be [1], the second one [2] and so on). This is the case of example of Figure 12.

- Bibliographic references can be sorted according to the surname of their first author (hence the first cite in the paper is not likely to be the [1]).

Moreover, specifying the author instead of using a number is also common in scientific writing. In **author-date** citations the surname of the author is used, followed by the year of publication. An example is (Ferrer 2005). If there is more than one author, there are several options: using only the first author, using all the authors if there are a few (Ferrer, Wu and Solanas 2005) or using the Latin expression "et al." (Ferrer et al.). Moreover, if the is more than one cite of the same authors and the same year, a letter is used to differentiate them. For example: (Ferrer and Solanas, 2005), (Ferrer and Solanas, 2005b).

However, in Information Technology disciplines it is more usual to find **summarised author-date** citations. In this case, there are some possible approaches: using a part of the surname of the first author, using the symbol '+' in case of several authors, etc. Some examples follow: [Ferr06], [Sol+94].

Finally, note that the style of citations usually depends on the style required by the publisher.

2.10 Acknowledgements

All the aforementioned sections are important since they are used for communicating the ideas and the results of a research work. However, research work is usually done under some kind of funding (that is, there is a government or organism which is paying for the research expenses, namely equipment, personnel, journeys, etc.).

On the other hand, some other people can collaborate in the paper but may not deserve appearing as authors. For instance, if a pair of colleagues have read the paper and have made some comments on it, their task should be acknowledged somehow.

Hence, the aim of the acknowledgements section is to thank funding organisms, anonymous reviewers, and personnel who have collaborated on the paper and deserve, at least, some gratitude.

Nonetheless, authors must be careful with acknowledgements. See this example:

> The author whishes to thank Agusti Solanas and Jordi Herrera for their assistance in preparing the previous work; Jordi Castellà for the drawings; Antoni Martínez Ballesté for programming the simulations and Josep Domingo for helping with writing the article.

The question is straightforward: which is the contribution of the author of the paper?

Finally, regarding financial support, the acknowledgements are sometimes placed in the first page of the paper, usually by means of a footnote in the title. This also depends on on specific requirements of the publisher.

2.11 Appendices

As a matter of fact, an article should be self-contained. This means that all the elements necessary to understand the research process and results must be written in the paper or, at least published in some other paper. Imagine the paper contains an expression whose understanding is not straightforward. Readers may need to know how did authors come up with the expression and, hence, they should be able to find the entire development of the expression somewhere.

In that sense, if the development of the expression has already been published, a citation to the paper must be enough. On the contrary, if the development has to be described in the paper and is too long to be inserted in the middle of the text, authors may consider putting it in an appendix.

Hence, the appendices include all the complementary information which is not put in the "body" of the paper for different reasons. In the body of the paper, authors should inform the reader about the information they can find in the appendices. For instance, authors could say "See Appendix 3 for the entire development of Equation 2".

3 Style of writing

When writing a paper, some considerations must be taken into account. Clearly, scientists are not expected to have the same writing capabilities as the most renowned writers have. As stated in the introduction of this module, the aims of scientific writing are clarity and brevity. In this section some general considerations and specific tricks to achieve these goals are given.

3.1 General considerations

There are some general considerations that authors must take into account for writing a scientific text. The next rules must be followed in order to success in elaborating a scientific text:

- A scientific text must be well-structured. You must divide the paper in sections (such those specified during this module), but sections should be divided into subsections, subsections may be divided into smaller parts… Which is the limit then? Clearly if you have a big section with a lot of text and this text elaborates on different topics, you could consider using a subsection for each topic.

- Your paper cannot be written in a few hours. You will start with a draft containing the main results and the main structure of the text. Then, after filling up each of the sections, you will have to read the paper several times and iterate through a read-and-review process. After the first version of the text is finished, it may be useful to let it be read by colleagues not involved with the text. Their comments are valuable since authors may omit some key aspects which are not considered because of the own knowledge of the authors.

- Include the necessary tables and figures which are useful for understanding the paper. If you are not a drawing specialist, ask someone for help.

Finally we must mention another important aspect: the mother tongue of the authors. Cleary, writing in English is likely to be easier for native English authors. However, having medium skills and reading lots of papers may be sufficient for writing scientific English. In addition, authors not having higher skills in English might need some help from colleagues or professional translators.

3.2 Language and numbers

To conclude, some specific considerations on writing are given. These basic rules are the linchpins of scientific writing.

Synonymy

In school writings, teachers often used to ask us using different words for a single concept. We had to demonstrate we had a rich vocabulary. However, using many synonyms in scientific writing is not a good practice. For example, if you are describing an electronic payment system composed of a server and a client, you should always make use of the word "client" instead of using client, user, purchaser, customer, and so on.

No ambiguity

Text should be proofread in order to assess there are no ambiguities. Here we have some funny examples:

- I know a man with a dog who has fleas (Who has the fleas: the dog or the man?)
- The robber was described as a tall man with a black moustache weighing 85 kilos (Of course, the man is assumed to weight 85 kilos).

Personal or impersonal style

In general, using either personal (e.g. We have designed a new protocol …) or impersonal/passive voice (e.g. A new protocol has been designed…) is possible in scientific writing. However, some publishers may specify that using passive voice is mandatory.

Hence, you can use either a personal or an impersonal style, but the fact is that the same style should be used all over the text.

Parallelism

The parallelism is the balance of style between the different parts of a sentence. The compound sentences are those with a certain risk of lack of parallelism. For instance:

> The algorithm has been implemented in assembly language and we noticed an important improvement in efficiency.

In this sentence, the first part uses the passive voice whereas the second one uses a personal style. A better writing for the sentence could be:

> The algorithm has been implemented in assembly language and consequently an important improvement in efficiency has been noticed.

Short sentences

Writing long sentences and using many relative clauses is not straightforward even for authors with advanced English skills. Moreover, long sentences are difficult to read. Since scientific writing must meet the clarity achievement, the authors could have split the sentence of the above example into two parts:

> The algorithm has been implemented in assembly language. Consequently, an important improvement in efficiency has been noticed.

Another way to reach clarity is by means of shortening sentences. For example:

> The improvement in the protocol is due to the suppression of step 7. Then a new implementation has been done. We finally observe that the new implementation clearly improves the results.

Could be shortened and written as:

> A new implementation, based on suppressing step 7, clearly improves the results.

Abbreviations and acronyms

In scientific and technical disciplines there is a plethora of large expressions which are usually named by their abbreviation. If you are about to use abbreviations in your text, you should consider the following hints:

- All the abbreviations should be introduced the first time they occur in the text.
- Naturally, there is no need for defining the well-known concepts (such as DNA, AIDS, laser, etc.).
- You can define an abbreviation if it has to be used several times during the text. For instance, if your paper deals with location-based services you can define the abbreviation LBS.
- Do not mix abbreviations and spelled-out terms throughout the text.

Gender

Handling gender of the subjects appearing in the text is somehow tricky. In English, you may refer to “the user”, which is the same whether the user is a man or a woman. However, if you later in the sentence want to use a pronoun, what will be the gender of the subject? Some authors prefer using the feminine gender. However this can be regarded as being discriminatory for the male gender:

> During the protocol the user sends a message. Then, she waits for the server response.

Some authors may use both genders at the same time:

> During the protocol the user sends a message. Then, he/she waits for the server response.

However, the best trick is to rewrite the sentence when possible so as have a plural subject:

> During the protocol, the users send a message. Then they wait for the server response.

Latin abbreviations

It is common that scientists use some Latin abbreviations throughout the paper. The most widely used are shown below:

- **e.g.** (exempli gratia), which means "for example".
- **i.e.** (id est), which means "in other words" or "that is".

Also note that **etc.** is one of the most popular common abbreviations (et cetera).

Numbers

When writing numbers, it is common to spell out numbers below 10:

> In the system there are three actors: the server, the clients and the trusted third party. The system can handle up to 15 client requests at the same time.

However, numbers must be spelled out when beginning a sentence.

Summary

This module is devoted to scientific writing. By studying this module, the basic skills for writing a research paper should be acquired.

During the first section, we introduce scientific writing and describe different kinds of scientific texts.

Throughout the second section, we elaborate on the different parts of a scientific paper. We show the importance of choosing the title and how tricky can be deciding the authorship. We focus on the elaboration of the abstract, since it might be published standalone. Moreover, we give some advice on elaborating the introduction, the conclusion of the paper, and the bibliography. We also give some hints on resources for describing the proposal of the paper and on how to show the results.

In the third section, we overview some basic rules that can be applied for successful writing of scientific text. We concentrate on language and grammar aspects.

Activities

1. During this module, use some Latin abbreviations. Can you list them along with its meaning?

2. Look for three scientific papers and identify the different parts. They should be related to three different disciplines: e.g. Computer Science, Chemistry and Medicine. Try to find differences between the structure and the style.

3. Make corrections to the following abstract. There are structure, style, grammar and spelling mistakes.

> The RFID technology is moving fast and it must be able to face the problems that its own growth implies. In [19] we presents a method for radio frequency identification. It was stated that scalability of the system was very bad. In this article, the protocol improves a lot.
>
> Our contribution is twofold: on the one hand, we propose a proposal in which the readers work together in order to perform the tags identification in a secure and scalable form.
>
> By appliying the proposed model, the scalability shortcomings of [19] are overcome and the complexity of the system does not grow exponentially and an implementation has been done.

4. Log to the IEEE Computer Society Digital Library and look for 5 abstracts. Elaborate a table to compare them so as to assess if they meet the basic requirements for an abstract: no references, present tense and summarising the whole paper.

Bibliography

Cargill, M., O'Connor, P. (2009) *Writing Scientific Research Articles: Strategy and Steps*, Wiley-Blackwell.

Day, R.A., Gastel, B. (2006) *How To Write and Publish a Scientific Paper* (6th. Ed.). Cambridge University Press.

Gustavii, B. (2008) *How to Write and Illustrate a Scientific Paper* (2nd ed.). Cambridge University Press.

Glossary

Abstract. Short summary of the paper (or another scientific text) that can be accessed standalone, separately of the body of the paper.

Acknowledgment. Some lines expressing the gratitude for those collaborating on the paper but not deserving appearing as authors. Also funding institutions are acknowledged in a scientific text.

Article (see paper)

Author. Anyone having worked hard on the elaboration of a scientific text. The only scientific text that must have only one author is the Ph.D. thesis.

Bibliography. Final section of a scientific text containing all the books and other references cited throughout the paper. It can also be called References.

Definition. Formal description of an element that takes part in a proposal.

Discussion. Section of the paper that elaborates on the achievement of the goals pointed out in the introduction and compares the proposal of the paper with those addressed in the background section.

Letter. Very short paper which usually discusses on previous works or presents proposals that do not need the extension of a paper.

Paper (or article). Scientific text that elaborates on a specific research process. It has a well-known structure and can be presented in a conference or published in a scientific journal.

Theorem. Strong mathematical statement which has to be broken into smaller statements in order to be proven.