Going through a paper in detail

I want to go through this paper in detail:

> *S. Fu, Y. Yu, M. Xu, A Secure Algorithm for Outsourcing Matrix Multiplication Computation in the Cloud, Proc. ACM Security in Cloud Computing, 2017.*
>
> dx.doi.org/10.1145/3055259.3055263

Why?

- It's a recent paper with Chinese authors.
- The research problem can be easily understood: (a) multiply two matrices together, (b) do it via cloud computing, and (c) do it securely.
- The paper is full of writing problems—the same problems I see in student papers.

Aim: Gain familiarity with the process of identifying and correcting writing problems.

This is all we're doing...

> *The basic problem in writing mathematics is the same as in writing biology, writing a novel, or writing directions for assembling a harpsichord: the problem is to communicate an idea. To do so, and to do it clearly, you must have something to say, and you must have someone to say it to, you must organize what you want to say, and you must arrange it in the order you want it said in,* **you must write it, rewrite it, and re-rewrite it several times, and you must be willing to think hard about and work hard on mechanical details such as diction, notation, and punctuation**. *That's all there is to it.*
>
> *— Halmos, How to write mathematics, 2009.*

Take home lessons

This is what I'm trying to get you to do...

1. We **identify useful information**, and **delete the rest**.

   A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, *for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts.* **This requires ... that every word tell.**

   — Strunk & White, Section III.13

2. There is no point in fixing the English in useless writing—just delete it instead.
   When there is nothing else but useful information, we then improve the fine details in the English writing.
3. When is information useful? That's tricky to answer, but I think basically... **when the reader learns something relevant**.

Going through a paper in detail

(front matter)

Title

A Secure Algorithm for Outsourcing Matrix Multiplication Computation in the Cloud

The advice we use here (and for every sentence):

> *A sentence should contain no unnecessary words...*
> — *Strunk & White, Section III.13*

Here's the main terms (the remaining words glue them together):

- Secure ✓
- Algorithm
- Outsource
- Matrix Multiplication ✓
- Computation
- Cloud ✓

Question: Which of these are **definitely necessary**?
Question: Can we throw away the rest? *Let's try!*

We try changing the title from:

A Secure Algorithm for Outsourcing Matrix Multiplication Computation in the Cloud

to:

Secure Cloud Matrix Multiplication

Question: Does the original title convey more information than this?
Yes! The original title implies "an algorithm" is presented. Let's fix this...

First, an algorithm is not actually presented in the paper; it's better described as a "protocol". However, the paper uses the word "scheme", so let's go with that.

A Secure Cloud Matrix Multiplication Scheme

Or variants:

A Secure Scheme for Cloud Matrix Multiplication

A Secure Scheme for Matrix Multiplication in the Cloud

Author names and affiliations

Shaojing Fu
College of Computer
National University of Defense Technology
Changsha,China
Sate Key Laboratory of Cryptology
Beijing,China
Science and Technology on Information Assurance Laboratory
Beijing,China

Yunpeng Yu*
College of Computer
National University of Defense Technology
Changsha,China
yuyunpeng@nudt.edu.cn

Ming Xu
College of Computer
National University of Defense Technology
Changsha,China

Problems:

1. We don't need to say department names three times: once is fine.
2. "College of Computer" is ungrammatical. (Do they only have one computer?) (Although, apparently this is what it's called—so we must use that, even though it's wrong.)
3. "Science and Technology on Information Assurance Laboratory" is ungrammatical. (Again, we must use that, even though it's wrong.)
4. "State" is misspelled "Sate"—unprofessional!
5. There's no space after the commas.

Shaojing Fu
College of Computer
National University of Defense Technology
Changsha,China
Sate Key Laboratory of Cryptology
Beijing,China
Science and Technology on Information Assurance Laboratory
Beijing,China

Yunpeng Yu*
College of Computer
National University of Defense Technology
Changsha,China
yuyunpeng@nudt.edu.cn

Ming Xu
College of Computer
National University of Defense Technology
Changsha,China

We change it to:

Shaojing Fu[1,2,3], Yunpeng Yu[1,†], Ming Xu[1]
[1] College of Computer, National University of Defense Technology, Changsha, China.
[2] State Key Laboratory of Cryptology, Beijing, China.
[3] Science and Technology on Information Assurance Laboratory, Beijing, China.
[†] Corresponding author. yuyunpeng@nudt.edu.cn.

Abstract

ABSTRACT

Matrix multiplication computation (MMC) is a common scientific and engineering computational task. But such computation involves enormous computing resources for large matrices, which is burdensome for the resource-limited clients. Cloud computing enables computational resource-limited clients to economically outsource such problems to the cloud server. However, outsourcing matrix multiplication to the cloud brings great security concerns and challenges since the matrices and their products often usually contains sensitive information. In a previous work, Lei et al. [1] proposed an algorithm for secure outsourcing MMC by using permutation matrix and the authors argued that it can achieve data privacy. In this paper, we first review the design of Lei's scheme and find a security vulnerability in their algorithm that it reveals the number of zero element in the input matrix to cloud server. Then we present a new verifiable, efficient, and privacy preserving algorithm for outsourcing MMC, which can protect the number privacy of zero elements in original matrices. Our algorithm builds on a series of carefully-designed pseudorandom matrices and well-designed privacy-preserving matrix transformation. Security analysis shows that our algorithm is practically-secure, and offers a higher level of privacy protection than the state-of-the-art algorithm.

Before even reading it, there's problems:

1. Words go outside the margins.
2. Large number of hyphenated words.
3. Weird line breaks. (The last line is just "m.")
4. The abstract contains a reference "[1]".
5. The spacing after "et al." is too big.

Abstract (sentence by sentence)

> Matrix multiplication computation (MMC) is a common scientific and engineering computational task.

Problems:

1. The reader knows what matrix multiplication is, and its importance.
2. Describing "matrix multiplication" as "matrix multiplication computation" confuses the reader: If the author meant "matrix multiplication", they would have written "matrix multiplication".
3. If we can avoid introducing an acronym (in this case MMC), we should.
4. Repetition: "computation" and "computational".

> *A sentence should contain no unnecessary words, a paragraph no unnecessary sentences.*
>
> — *Strunk & White, Section III.13*

We need "Matrix multiplication"; we can get rid of the rest.

Seinfeld (US TV series)

Sid: *Well I'm going down to visit my sister in Virginia next Wednesday, for a week, so I can't park it.*

Jerry: *This Wednesday?*

Sid: *No, next Wednesday, week after this Wednesday.*

Jerry: *But the Wednesday two days from now is the next Wednesday.*

Sid: *If I meant this Wednesday, I would have said this Wednesday. It's the week after this Wednesday.*

Reader's logic:

> *If the author meant X, the author would have said X. Since the author did not say X, the author therefore does not mean X.*

Moral of the story: If there is an obvious way to say X, we should say it that way. Otherwise the reader will think we mean something else.

Next sentence:

> But such computation involves enormous computing resources for large matrices, which is burdensome for the resource-limited clients.

Problems:

1. Writing "such computation" implies we're using the context from the previous sentence. Why not just write one sentence instead?
2. What does "enormous computing resources" mean? Vagueness. And "enormous" usually refers to physical size.
3. How big are "large matrices"? Vagueness.
4. What is a "resource-limited client"? Vagueness. (Are there non-resource-limited clients too?)
5. Why are we talking about "clients" all of a sudden?
6. Grammar quibble: "... for the resource-limited clients" should be "... for resource-limited clients".

We now edit these two sentences:

> Matrix multiplication computation (MMC) is a common scientific and engineering computational task. But such computation involves enormous computing resources for large matrices, which is burdensome for the resource-limited clients.

We change this to:

> A range of science and engineering applications require multiplying matrices with perhaps millions of rows and columns, requiring significant computational resources.

It's still not perfect:

- I'm just guessing "millions of rows and columns"; I would need to investigate deeper to check this is actually correct.
- It still has the vague "significant computational resources"; again, I would need to investigate deeper to figure out what this means.

To improve our writing, we need to be familiar with the computer science.

Next sentence:

> Cloud computing enables computational resource-limited clients to economically outsource such problems to the cloud server.

Problems:

1. The conference is called *Security in Cloud Computing*; the audience knows what cloud computing is—**delete this sentence!!**
2. What is a "computational resource-limited client"? Vagueness. (Are clients who aren't "computationally resource limited" unable to use cloud computing?)
3. What does "economically outsource" mean? Vagueness. This also suggests the paper is about monetary costs—it's not!
4. What does "such problems" mean? Vagueness.
5. What's the difference between "cloud server" and "cloud"?

> However, outsourcing matrix multiplication to the cloud brings great security concerns and challenges since the matrices and their products often usually contains sensitive information.

Problems:

1. Why write "great security concerns and challenges"? Salesmanship.
2. What's the difference between a "security concern" and a "security challenge"? Vagueness.
3. I think "often usually" is a typing error. Need to proofread.
4. Grammar quibble: "contains" should be "contain".
5. I find it hard to believe that matrices that need multiplying together "... usually contain sensitive information."
6. A general writing guideline is that the most important information goes at the end of the sentence.

> However, matrices and their products may contain sensitive information, so outsourcing matrix multiplication to the cloud brings security concerns.

> In a previous work, Lei et al. [1] proposed an algorithm for secure outsourcing MMC by using permutation matrix and the authors argued that it can achieve data privacy.

Problems:

1. Grammar quibble: "In a previous work ..." should be "In previous work ..." or "In prior work ...". Why even say this? (The word "proposed" is past tense.)
2. LaTeX interprets the full stop in "Lei et al." as the end of the sentence. It should be typeset:
   (a) `Lei et al.\ \cite{...}` or
   (b) `Lei et al.~\cite{...}`.
3. We avoid citations in abstracts, as abstracts appear outside of paper (e.g., on webpages). Instead we write e.g. "Lei et al. (2014)".
4. Grammar quibble: "secure outsourcing" should be "securely outsourcing".
5. Grammar quibble: "by using permutation matrix" should be "by using permutation matrices".
6. How on Earth does a permutation matrix allow one to outsource matrix multiplication? It doesn't make sense.
7. Lei et al. (2014) does not use permutation matrices. Error.
8. What does "the authors argued" mean? Vagueness.
9. Repetition: "secure" and "achieve data privacy".

Side note...

We found an error in the first few sentences of Fu et al. (2017).

We can be confident that the paper contains many more errors.

Low-quality writing and errors go hand in hand:

- low-quality writing make errors hard to find;
- low-quality writing indicates a lack of proofreading (where the author would find the errors); and
- if the authors don't care about one ("writing quality" or "errors"), they probably don't care about the other. In fact, they probably don't care about the paper at all—they just want it published.

When I see a poorly written paper, I think:

The paper is poorly written.

Therefore, the paper is probably hard to read.

And the paper probably contains errors. (Or is total rubbish.)

Therefore, it's probably a waste of time reading it.

We now edit this:

> In a previous work, Lei et al. [1] proposed an algorithm for secure outsourcing MMC by using permutation matrix and the authors argued that it can achieve data privacy.

First, we download Lei et al. (2014)[1] and find out what they actually did:

> *The main idea to protect the privacy is employing some transformations on the original MMC problem to get an encrypted MMC problem which is sent to the cloud; and then transforming the result returned from the cloud to get the correct result to the original MMC problem.*
> — *Lei et al., ... outsourcing large matrix multiplication ..., Inf. Sci., 2014.*

Thus we change the sentence to:

> Lei et al. (2014) described a matrix transformation protocol for securely outsourcing matrix multiplication.

[1] doi.org/10.1016/j.ins.2014.05.014

> In this paper, we first review the design of Lei's scheme and find a security vulnerability in their algorithm that it reveals the number of zero element in the input matrix to cloud server.

Problems:

1. Writing "In this paper, ..." is often unnecessary.
2. Error: "Lei's scheme" should be "the scheme by Lei et al." ("et al." means "and others"). Some people write "Lei et al.'s scheme".
3. Why write "we first review the design of Lei's scheme"? Is this not obvious?
4. Grammar quibble: "to cloud server" should be "to the cloud server".
5. When multiplying matrices X and Y, which matrix is "the input matrix"? Vagueness.

We now edit this:

> We find a security vulnerability in the scheme by Lei et al.; it reveals the number of zeros in the input matrices.

(Note: Both "zeros" and "zeroes" are fine, as long as we're consistent.)

At this point, the reader will probably think: *How is revealing the number of zeros a security vulnerability?*

The next sentence should address this. E.g. we might write:

> *This vulnerability is significant e.g. when [blah].*

The authors do not write such a sentence. This makes me feel like it's not a meaningful security vulnerability.

Another comparable "vulnerability": Lei et al. (2014) also reveals the dimensions of the input matrices. Just like the scheme by Fu et al. (2017).

The reader has a similar thought process to before:

> The paper is poorly written.
>
> Therefore, the paper is probably hard to read.
>
> And the results are probably insignificant.
>
> Therefore, it's probably a waste of time reading it.

> Then we present a new verifiable, efficient, and privacy preserving algorithm for outsourcing MMC, which can protect the number privacy of zero elements in original matrices.

Problems:

1. Unnecessary word: "Then".
2. Grammar quibble: "a new verifiable, efficient, and privacy preserving algorithm" should be "a new verifiable, efficient, privacy-preserving algorithm". Two bugs:
   (a) we don't put "and" for a list of adjectives before a noun (e.g. "big round black ball" not "big, round, and black ball"), and
   (b) "privacy-preserving" is used as an adjective (a compound adjective) which requires a hyphen ("-").
3. It implies the method by Lei et al. (2014) is not verifiable, efficient, nor privacy-preserving. Misleading.
4. What does "which can protect" mean? (Does it actually protect, or not?) Vagueness.
5. Grammar **disaster**: "protect the number privacy of zero elements".
6. Grammar quibble: "in original matrices" should be "in the original matrices".

Next sentence:

> Our algorithm builds on a series of carefully-designed pseudorandom matrices and well-designed privacy-preserving matrix transformation.

Problems:

1. "The proposed algorithm" is more impartial than "Our algorithm". (Think: "it's ours, not yours".)
2. What's proposed is not really an algorithm.
3. Why write "carefully-designed" and "well-designed"? Salesmanship.
4. What does it mean to "build on a series of ... matrices"? Vagueness.
5. Grammar quibble: "carefully-designed" should be "carefully designed" (an exception to compound adjectives is when the first word is an adverb).
6. Grammar quibble: "... and well-designed privacy-preserving matrix transformation" should be "... and a well-designed privacy-preserving matrix transformation".
7. Does "privacy-preserving" simply mean "we don't give away the number of zeros"? Misleading.
8. What is the "matrix transformation"? Vagueness.

We now edit these two sentences:

> Then we present a new verifiable, efficient, and privacy preserving algorithm for outsourcing MMC, which can protect the number privacy of zero elements in original matrices. Our algorithm builds on a series of carefully-designed pseudorandom matrices and well-designed privacy-preserving matrix transformation.

We change this to:

> We instead propose a scheme which first adds noise to the input matrices, then multiplies the resulting matrices in the cloud, then subtracts the post-multiplication noise from the product.

> Security analysis shows that our algorithm is practically-secure, and offers a higher level of privacy protection than the state-of-the-art algorithm.

Problems:

1. What does "Security analysis" mean? Vagueness.
2. Better to write "the proposed scheme" instead of "our algorithm".
3. Grammar quibble: "practically secure" not "practically-secure".
4. Why write "offers a higher level of privacy protection"? Just say what it actually does!!! Vagueness.
5. Saying "a higher level of privacy protection" for simply concealing the number of zeros is overselling it—it's not that important! Salesmanship.
6. Which algorithm is "the state-of-the-art algorithm"? Vagueness.
7. Reading the paper, we find that we get the security of Lei et al.'s scheme, only if we use it to multiply the matrices. Misleading.

Thus we change this sentence to:

> We verify mathematically that the proposed scheme does not reveal the number of zeros in the input matrices. Moreover, the proposed scheme also inherits the security properties of the scheme by Lei et al., if we use it to multiply matrices in the cloud.

Experimental results

The authors also don't mention their experimental results in the abstract. Ordinarily this would be a huge mistake. But...

- In their experiments, the proposed scheme performs matrix multiplication nearly three times as fast as the scheme by Lei et al.
- This implies they did not use the method by Lei et al. to multiply matrices.
- This implies that, in their experiments, the privacy is less than that of Lei et al. Misleading.

I wouldn't even include these experimental results in the paper—they're deceptive.

The experiments should be redone using the method by Lei et al. to multiply matrices in the cloud. The experimental results should then demonstrate that the additional overhead is not too large.

Again, we see that proper writing highlights significant problems (beyond the grammar).

New abstract (draft)

Putting those together:

> A range of science and engineering applications require multiplying matrices with perhaps millions of rows and columns, requiring significant computational resources. However, matrices and their products may contain sensitive information, so outsourcing matrix multiplication to the cloud brings security concerns.
>
> Lei et al. (2014) described a matrix transformation protocol for securely outsourcing matrix multiplication. We find a security vulnerability in the scheme by Lei et al.; it reveals the number of zeros in the input matrices. We instead propose a scheme which first adds noise to the input matrices, then multiplies the resulting matrices in the cloud, then subtracts the post-multiplication noise from the product. We verify mathematically that the proposed scheme does not reveal the number of zeros in the input matrices. Moreover, the proposed scheme also inherits the security properties of the scheme by Lei et al., if we use it to multiply matrices in the cloud.

We can still make it read better as a whole.

- "Lei et al." is repeated three times, and we can reduce this by rephrasing the text in blue.

New abstract (after polishing)

> A range of science and engineering applications require multiplying matrices with perhaps millions of rows and columns, requiring significant computational resources. However, matrices and their products may contain sensitive information, so outsourcing matrix multiplication to the cloud brings security concerns.
>
> Lei et al. (2014) described a matrix transformation protocol for securely outsourcing matrix multiplication; we find a security vulnerability in this scheme, namely, it reveals the number of zeros in the input matrices. We instead propose a scheme which first adds noise to the input matrices, then multiplies the resulting matrices in the cloud, then subtracts the post-multiplication noise from the product. We verify mathematically that the proposed scheme does not reveal the number of zeros in the input matrices. Moreover, the proposed scheme inherits the security properties of the scheme by Lei et al., if we use it to multiply matrices in the cloud.

It still needs an explanation as to (a) why revealing the number of zeros is important, (b) experimental results [for suitable experiments], and so on. (But I can't do that.)

Note: Other conferences might need a sentence to explain "cloud computing".

Going through a paper in detail

(introduction)

Introduction

First sentence:

> With cloud computing becoming more widely utilized, more and more users with computational resource-constraint devices tend to outsource their computing needs to the cloud server which has plenty of computing resources in a pay-per-use manner, relieving the clients from computation burden.

What useful information do we learn from this?

1. An increasing number of users are using cloud computing.
2. We can outsource computation to the cloud.
3. We pay per use. (Bear in mind the paper does not talk about pricing.)

What does the reader need from this to understand the paper? I think there's nothing useful (for a conference called *Security in Cloud Computing*). The sentence is a waste of the reader's time—delete.

Second sentence:

> However, directly outsourcing computation to the cloud inevitably brings in new security concerns and challenges[2, 3].

What useful information do we learn from this?

1. Using the cloud brings (unspecified) security concerns.
2. We can see references [2] and [3] to learn about security concerns in cloud computing.

Third sentence:

> The first concern is the *privacy* of input and output data.

What useful information do we learn from this?

1. One type of security concern we're interested in is privacy.
2. It relates to "input and output data", which I think should say "user data". (But this is obvious—what else could it relate to?)

Fourth sentence:

> Data privacy has become a critical issue for cloud users when they host their data on remote and untrusted cloud storage.

What useful information do we learn from this?

1. The privacy concerns are "critical", whatever that means...
2. Cloud storage is "remote and untrusted". (Although, this is obvious.)

The first four sentences are:

> With cloud computing becoming more widely utilized, more and more users with computational resource-constraint devices tend to outsource their computing needs to the cloud server which has plenty of computing resources in a pay-per-use manner, relieving the clients from computation burden. However, directly outsourcing computation to the cloud inevitably brings in new security concerns and challenges[2, 3]. The first concern is the *privacy* of input and output data. Data privacy has become a critical issue for cloud users when they host their data on remote and untrusted cloud storage.

And we merge those four sentences to:

> Utilizing the cloud for outsourcing computation brings important data *privacy* concerns [2, 3].

Importantly, **we save the reader's time**.

> To protect the privacy of sensitive data, the client should encrypt the sensitive data before outsourcing and decrypt the returned results from the cloud after outsourcing.

This sentence is almost all useful information! But...

1. Why write "should encrypt" instead of just "encrypts"? Vagueness.
2. Grammar quibble: "sensitive data" instead of "the sensitive data".
3. We don't need to say "... from the cloud after outsourcing"; it's clear from context.
4. It says "sensitive data" twice.

We change this to:

> To ensure privacy, the client encrypts sensitive data before outsourcing, and decrypts the returned results.

It sounds better to say "safeguard privacy" or "ensure privacy", than "protect privacy".

Next snippet:

> The second concern is the *verification* of the outsourcing computation results. The cloud is not fully trusted. For example, for the financial incentives, the cloud may decrease the amount of the computations and then return invalid results. Consequently, the client needs to verify the correctness of the returned outputs.

We perform the same process...

> A second concern is *verification* of the correctness of the results returned from the cloud.

Next snippet:

> The third concern is *efficiency* of outsourcing computation. In the outsourcing process, the computation on the client side must be substantially smaller than performing the original computational problem on its own.

Again we perform the same process...

> A third concern is the *efficiency* of outsourcing computation; in order for cloud computing to be worthwhile, the client-side computation must be substantially less than the cloud's computation.

Here, I add "in order for cloud computing to be worthwhile" to explain why we "must" have this property.