

张程 2013310558

2013年12月25日

## Introduction

In this paper analysis I choose four paper from Martin Casado who is a PhD student from Stanford University majors in Computer Networks. These four papers created the era of SDN and attracted lots of people to do their research in this field.

The four papers are listed below :

- Virtualizing the network forwarding plane. Martin Casado. 2010
- SANE: a protection architecture for enterprise networks. Martin Casado. 2006
- Fabirc : a retrospective on evolving SDN. Martin Casado. 2012
- Rethinking enterprise network control. Martin Casado. 2009

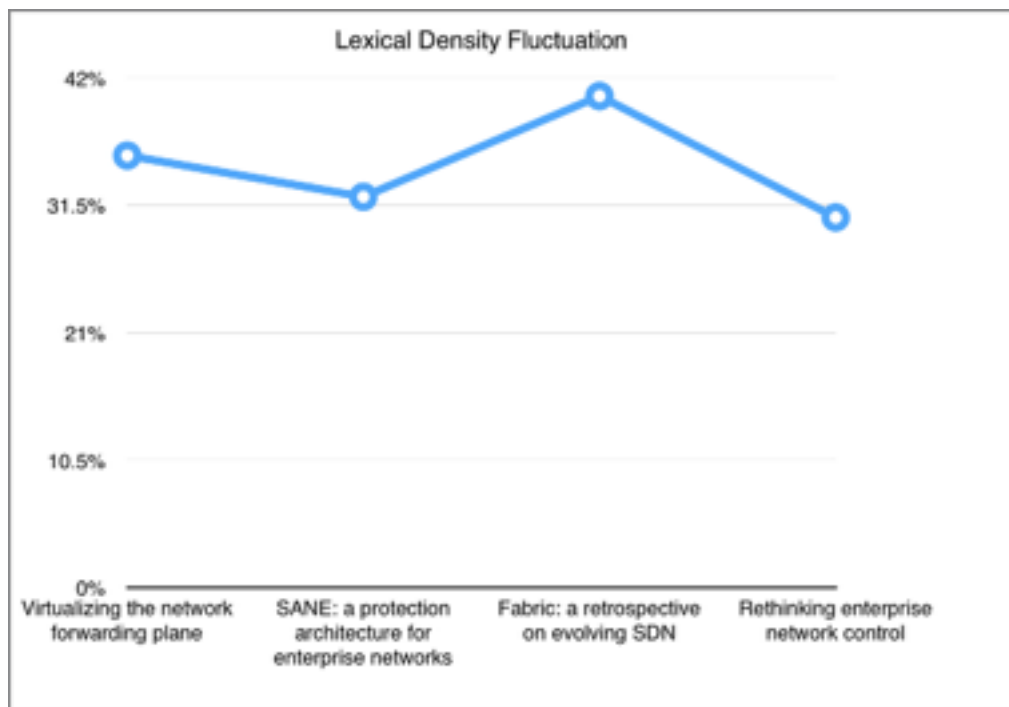
The main focus of this report is to list useful words for my further study, also I want to see wether the verbs and adjectives which the author used has a fixed pattern which can be used to identify the author.

## Words & Sentence Analysis

I first make the words and sentence analyzing, the results are following.

	Virtualizing the network forwarding plane	SANE: a protection architecture for enterprise networks	Fabirc : a retrospective on evolving SDN	Rethinking enterprise network control
Total Words Count	3158	6807	3043	7964
Number of Different Words	1124	2191	1233	2433
Lexical Density	35.6%	32.2%	40.5%	30.5%
Sentence Count	412	1108	525	1203
Average Sentence Length	12.75	10.59	10.53	11.53
Max Sentence Length	65	44	54	59

figure 1 words and sentence count



From the data above, we can conclude that the average Lexical Density is rather firm, it ranges from 30.5% to 40.5%, and the Average Sentence Length also has a relatively fixed length about 11 words per sentence. I draw the Lexical Density figure to show the tendency. In academic paper, the number of words in a sentence should be readable, so that 11-13 words per sentence is the moderate number. From the figure above we can also see that the number of different words declines with the total words growing.

## Top Frequency Words & Key Words

In this section I choose five top used words from each paper and I made a experiment to see whether these five words can lead the google search engine to find the specific paper. According to well-known knowledge, the key words of a paper can be used by search engine like Google to find the paper, but what about the top frequency words in a paper? Are they and key words the same word?

I select top five occurrence word in each paper, then put them into search engine, check the position of the right link which lead to the paper.

SANE: a protection architecture for enterprise networks	Occurrences	Frequency
Network	117	1.7%
Sane	97	1.4%
Switches	61	0.9
Capability	61	0.9
End	61	0.9

Virtualizing the network forwarding plane	Occurrences	Frequency
Logical	132	4.2%
Physical	110	3.5%
Network	99	3.2%
Forwarding	97	3.1%
Plane	37	1.2%

Fabirc : a retrospective on evolving SDN	Occurrences	Frequency
Network	81	2.7%
Fabric	62	2%
Edge	54	1.8
Forwarding	35	1.2
Packet	35	1.2

Rethinking enterprise network control	Occurrences	Frequency
Network	149	1.9%
Controller	124	1.6%
Switch	114	1.4%
Ethane	114	1.4%
Flow	103	1.3%

After I enter the high frequency words from the result into google engine, the result is this :

	Match Item Position from Google
SANE: a protection architecture for enterprise networks	1
Virtualizing the network forwarding plane	1
Fabirc : a retrospective on evolving SDN	1
Rethinking enterprise network control	12

The reason for paper: Rethinking Enterprise Network Control' first match item falling behind is that while other papers' key words also in the high frequency words list, this paper's high frequency words do not show the specific title. From above we can conclude that in these papers we experiment, the frequency of a word has nothing to do with the title or the key words. Key words often can lead us to the specific paper in search engine cause they show the main thought of that paper.

## Verbs

In this section I select a list of report verbs to test four papers, the report verbs are :

describe	study
find	develop
report	propose
show	use
suggest	demonstrate
observe	analyze

I count every report verbs in each paper to see wether there is a fixed pattern of the writer of paper. The result is below:

	Virtualizing the network forwarding plane	Rethinking enterprise network control	Fabirc : a retrospective on evolving SDN	SANE: a protection architecture for enterprise networks
describe	11	15	5	3
find	0	8	2	2
report	0	1	0	0
show	0	10	0	9
suggest	0	9	0	1
observe	0	1	1	0
study	0	3	0	4
develop	0	5	1	4
propose	1	3	4	0
use	13	152	32	101
demonstrate	0	0	0	3
analyze	0	0	0	1

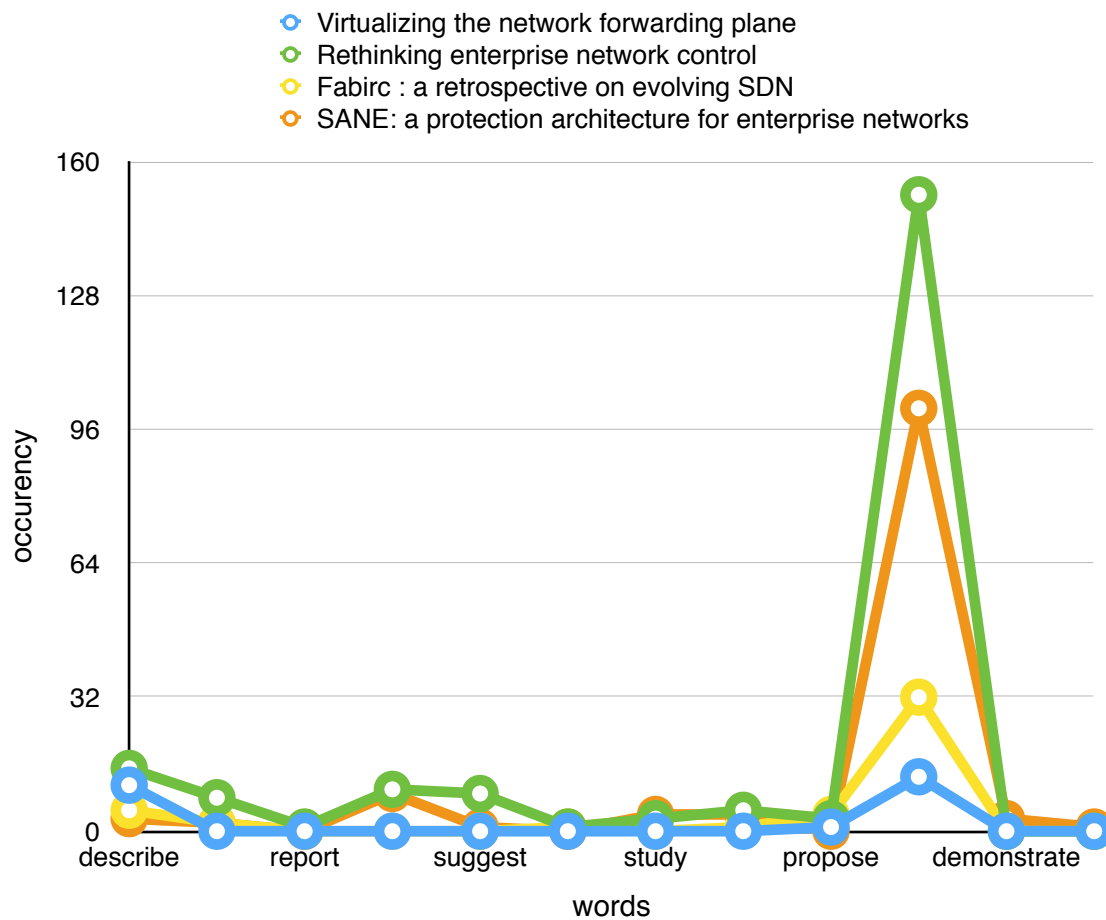
From the figure below we can conclude these :

- there is a kind of fixed pattern in writer's paper
- the writer pattern can change over times
- one author has his own favorite words, for Martin is "use", "describe", "find", his unfavorite words are "demonstrate", "analyze", "observe", "report".

From the data we can also see the word like "show", it appears in one paper for about 10 times, but doesn't appear in another paper, I assume the reasons are:

- some paper was written in early years of the writer's PhD life, the writing word has changed since the writer were getting more academic.

- some words like show are content specified, a paper with more figures will has more show words.



## Adjectives Words

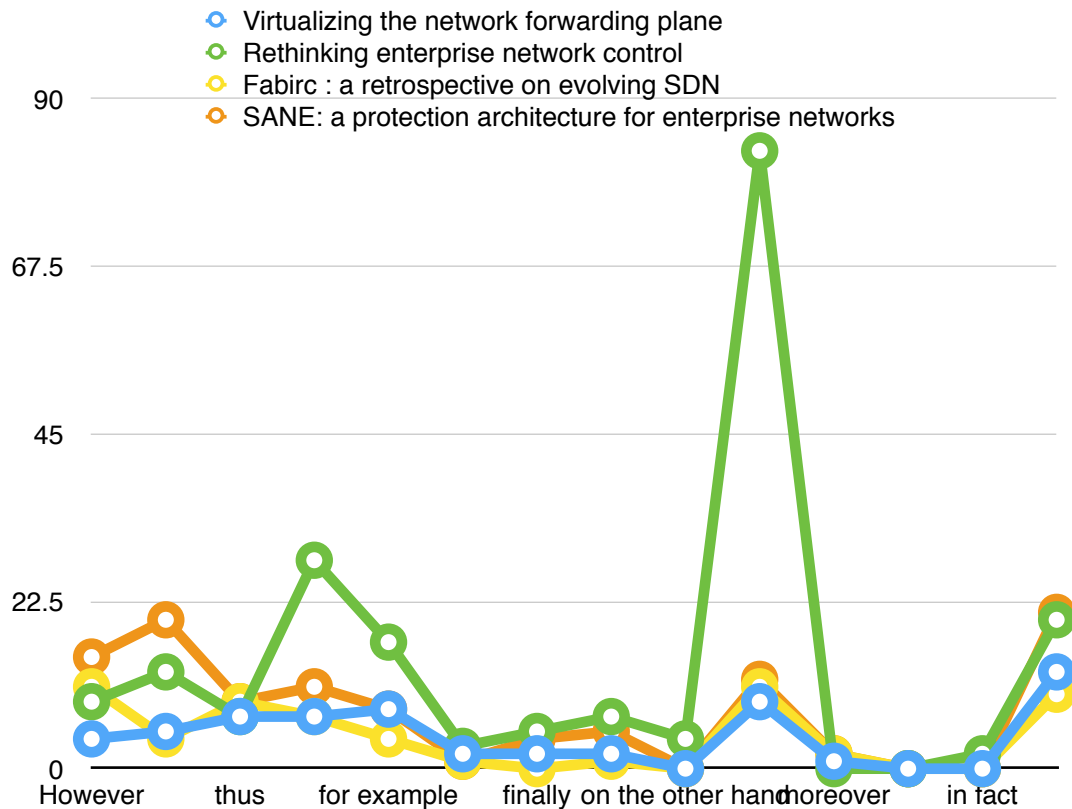
In this section I choose the common adjective words to see wether there is a fixed pattern in author's writing. The adjective words are :

However	finally
first,second	therefore
thus	on the other hand
also	then
for example	moreover
in addition	nevertheless
such as	in fact

	Virtualizing the network forwarding plane	Rethinking enterprise network control	Fabirc : a retrospective on evolving SDN	SANE: a protection architecture for enterprise networks
However	4	9	11	15
first,second	5	13	4	20
thus	7	7	9	9
also	7	28	7	11
for example	8	17	4	8
in addition	2	3	1	1
finally	2	5	0	4
therefore	2	7	1	5
on the other hand	0	4	0	0
then	9	83	11	12
moreover	1	0	2	2
nevertheless	0	0	0	0
in fact	0	2	0	0
such as	13	20	10	21

From the data above we can see the adjective words using is relatively fixed for the author, however, the most strange point is the paper : Rethinking Enterprise Network Control. In this paper the adjective words : “for example” and “ also” “ in fact” are low frequency in other papers but high occurrence in this paper. This paper is written in 2009,it is the second old paper in four papers, and it show a uncommon way of the author’s writing. Based on above data, i doubt wether this paper is writer himself wrote? or cooperate with someone else.

From the figure above, we can see the strange behavior of the green line : Rethinking enterprise network control.



## conclusion

I summarize the useful verbs and adjective words from the paper. Also I come the conclusion that there is a fixed pattern between pattern and author, the pattern to some degree represent the author's writing habit. But the pattern can change over times, as a person becoming more and more specialized in the research field, the academic words he use will change, but the basic verbs like "find", "use" will be used as ever.

describe	propose
find	use
report	show
However	finally
first,second	therefore
also	in addition
for example	moreover
such as	



作业思考：通过做这个作业，我觉的自己首先学会了以后在看论文的同时还要注意他的用词和一些链接表达，并且把自己遇到的好的表达方式标记下来。另外一点是，在做作业的过程中我也学习了一些spss中有关词语分析的软件，我觉的一个作者的写作语言和常用句型结构应该是相对固定的，那是不是可以利用现有的数据分析软件分析获得一些著名作者的写作模式和常用句型呢。我觉的这些模型是否可以应用到一些英文论文重复度或者抄袭检测系统中。