University of Oulu
Faculty of Science
Dpt. Inf. Proc. Sci.

UNIVERSITY OF OULU
Department of Information
Processing Science

Opponent’s opinion of the Master Thesis
’The Stream Cipher RC4’
by Michal Hojsik

In this thesis the secret inner states of RC4 are studied. Ever since the replacement of DES by AES as an encryption standard, RC4 has been the most popular secret key cipher in general use. Encryption is one necessary and indispensable key to secure communication over a hostile environment (like Internet) nowadays. This is why in the information security scientific community much attention is paid both to the design of new ciphers and to attacks against and analysis of the classical ones to reveal their possible hidden vulnerabilities. It is remarkable that the attack and analysis research very many times is carried out, not by computer scientists, as one easily could expect, but by mathematicians through the detection of the inner structures of ciphers. This is also the subject matter of the thesis at hand. Hojsik presents new and creatively developed methods to inspect the RC4 encryption system. Stepwise, by the

1. careful analysis of persistent states of RC4
2. construction of the tabular model
3. matrix representation of any equivalences on a finite linearly ordered set

an efficient toolbox to analyze the inner states of RC4 is constructed. The author proves the periodicity of persistent table triples and, in the latter part of the thesis, the nonexistence of persistent states in several special cases. Finally, the significance of persistent states to the security of RC4 is considered. The basic problem concerning the existence of persistent states remains unsolved offering a challenge for further research. In the work there exist a lot of brilliant ideas and considerations, a clear proof of the author’s ability to scientific work. The results achieved, after small further processing, are certainly worth publishing. The thesis is written, except some minor misprints, in excellent English. On the other hand, the reader catches the impression that in the formal side something is missing, left half-done. Especially the mathematical manner of representation leaves quite a lot to hope. The definitions are often loosely written, one of them even in a deductive way implying that a concept is used to define itself. The purpose of a definition is to give a name and/or structure to a concept, nothing else. A definition is not a saga, it is not a place to declare results. It is the supervisor's assignment to guide the student in this respect.

A list of small omissions and misprints found in the thesis as well as some personal comments can be found below.

Altogether, Michal Hojsik has done fine work and I sincerely recommend that the thesis is accepted to its purpose with grade 2 (very good).

In Oulu on Tuesday, May 23rd 2006

[redacted]

Juha Kortelainen
Professor

UNIVERSITY OF OULU
Department of Information
Processing Science

The detailed comments, questions and suggestions are as follows.

Page 1: The name of the thesisi could be 'On the analysis of the stream cipher RC4'.

Page 3, line +6: 'work' instead of 'Work'.

Page 3, line +9: 'A closer look' instead of 'Closer look'.

Page 4, line +2: The name of the author is misspelled.

Page 5 line +4: The name of the supervisor is misspelled.

Page 5: A more extensive Introduction could be in order.

Page 6, line +18: What does the sentence 'stream ciphers have limited or no error propagation' mean?

Page 7, line +16: The set of states is normally denoted by S, a sparate state by small s (with or without a subscript).

Page 10, line -13: Some notation about permutation groups could be included.

Page 11, line +7: The parameter k is not defined.

Page 12, line -15: 'occurs' instead of 'occur'.

Page 12, line -13: in 's-box' the letter s in italics.

Page 12, line -13: 'once' instead of 'one'.

Page 15, line +3: Something is wrong in "the IV is used IV'

Page 15, line +11: 'The paper' instead of 'Paper'.

Page 16, lines +14: 'The author' instead of 'Author'.

Page 16, line -12: 'the author' instead of 'author'.

Page 17, Section 4.1: Definition 1 is clumsy. The parameters i, j and σ should be given in the definition, not before. The last sentence is a remark that should be situated after the definition.

Page 17, line -10: 'In terms of' instead of 'In the terms of'.

Page 19, the title of Section 4.2: 'A closer look' instead of 'Closer look'.

Page 19: Notation is clumsily written. If you give a set in three dot notation, at least two first elements should be writen down so that the reader can find out the rule: 'let $k \in \{a, a+1, \ldots, h\}$ ... The mistake recurs in several places.

Page 19, Definition 8: The last sentence does not belong to the definition at all.

Page 19, Definition 10: Let $T_0 = (i_0, j_0, \sigma_0)$ be a persistent state. For each $a \in \Sigma$ and $k \in \mathbb{N}$ let $\eta_k(a) = k' - k$ where k' is the smallest integer $r > k$ such that $\sigma_r(j_r) = a$. Let $\eta_k(a) = \infty$ if such an r does not exist.

Page 20, line +17: You have to fix a and k in the proof since this is not done in the proposition.

Page 20, Definition 12: 'is reduced' instead of 'is said to be reduced'. The mistake recurs in several places.

Page 20, lines -2 and -1: 'and let $a \in \Sigma$ be such that $\sigma_k(j) = a$.' instead of 'and set $a = \sigma_k(j)$.'

Page 21, line +3: Use, for instance, the symbol b instead of a (the latter already fixed).

Lemma 13 can easily be proved directly which is more comfortable to the reader to follow.

Page 21, Lemma 15: The proof is clumsy. Where is the induction assumption applied?

Page 22, line +14: 'We then have' instead of 'we have'.

Page 22, lines -15 and -16: I do not understand the contents of the sentence on these lines.

Page 23, Definitions 19 and 20: The two definitions form together a tale. They should be unified and the process described and developed in normal text, not in a definition. Then phrases like 'we say that' (when concepts are defined) as well as small remarks and implications could be applied. By the way, v and t_0 should be mentioned to be functions.

Page 24, line +9: t_{l+1} instead of t_l.

Page 24, line +13: ', v and the value' instead of 'and the value'. The words 'uniquely determine' should be chosen otherwise or explained since t_{i+1} and μ_{i+1} are determined by t_i and μ_i in the definition.

The last sentence of the proof of Lemma 23 should be the first sentence of it.

Page 24, line +20: What does the expression 'adjective table' mean?

Page 24, line -9: 'highlighted, as well as the value in the first row and μ_1-st column' instead of 'highlighted.'

Page 24, line -6: 'column' instead of 'row'.

Page 28, row +1: In Notation, the restriction $(v^{(-a)}, t_0^{(-a)}, \mu_0)_N^n$ of a table triple is (by definition) not any more a table triple since the functions v and t_0 are not total any more.

Definition 28 is deductive and difficult for the reader to catch. Do the sets run empty or do we stop when the size is one? The idea in the definition should be explained in advance.

Lemma 29, proof, part a) The symbols a and b are used as if they where fixed. The existential quantification of them is, however, not enough to fix the symbols.

Page 28, line -7: 'finitely' instead of 'finally'.The reasoning for the case (i) should be more rigorous.

Page 29, line +7: 'finitely many persistent' instead of 'finally'.

Page 29, line +8: Erase the sentence 'which are persistent'.

Page 29, line +15 and +16: 'contradiction' instead of 'the contradiction'. Reasoning in the parenthesis should be more accurate, Lemma 23 should be mentioned.

Page 31: Definitions 35 an 36 are clumsy; one can just fix the persistent state to be (v, t_0, μ_0) and then, in the normal text, define the respective concepts. Why do you use capital P instead of small p? Capitals are for sets, small letters for numbers.

The notation should be corrected as instructed above through the thesis.