University of Zurich[UZH]

**Department of Informatics**

University of Zürich
Binzmühlestrasse 14
CH-8050 Zürich
Switzerland

**Prof. Dr. Burkhard Stiller**
Communication Systems Group
CSG@IfI
Phone: +41 44 635 6710 (direct)
+41 44 635 4569 (Secretary)
Fax: +41 44 635 6809
E-mail: stiller@ifi.uzh.ch
URL: http://www.csg.uzh.ch/

***To Whom It May Concern***

Zürich, September 10, 2018

**Letter of Recommendation for Mr. Tomáš Cejka's Doctoral Thesis on "Stream-wise Parallel Anomaly Detection in Computer Networks"**

Dear Sirs,

the thesis of Tomáš Cejka on "Stream-wise Parallel Anomaly Detection in Computer Networks" has addressed a very timely and relevant topic in the areas of today's networking sector, especially within the areas of a flow-based network monitoring and network security. Since a stream-wise processing determines a suitable principle for performing a security analysis of large-scale computer networks' data in transmission, all flow records have to be processed on-the-fly upon reaching the flow collector. Thus, the NEMEA framework had been developed and investigated in that respect as a proof-of-concept.

Tomáš shows that such an approach is possible and provides for a technically feasible approach, including the addressing of a key parameter of ever growing traffic based on parallel flow-based analysis. Major considerations have been evaluated and experimented with, where these had been performed on data sets from real backbone traffic of the Czech National Educational Network. Thus, Tomáš defines and investigates in this research work in detail his approach proposed to enforce a practically deployable solution for parallel anomaly detection and ensures that available resources are effectively controlled for a parallel use. Such an approach is new and in goes well beyond major related work.

**Content and Contributions**

Section 1 introduces the major basics of the thesis and summarizes the motivation, defines the targeted at challenges, and lists the thesis goals. This is carefully phrased into the problem statement, which determines as the challenging part the distribution of a single stream of data among multiple nodes, whereas each node should process just a subset of the original stream. Thus, a suitable mechanism of data distribution must be found, which preserves all results of a possible detection. Furthermore, since traditional flow records with very basic information are not sufficient for a detection of advanced threats — they may be similar to legitimate traffic —, the reliability of the detection algorithm needs to be maintained for application layer traffic (L7). To address these aspects, an open source tool, due to lacking any suitable elements as of today, was determined and proposed to be developed.

Section 2 describes the current state-of-the-art and lists briefly related work in the area of network security, anomaly detection, and parallel processing using flow data. While monitoring approaches are discussed, relevant detection methods are summarized, too. The “Big Data Processing” is focussed on parallel processing, while adding semantics of data as a discussion item. Thus, Section 2 does address the key input factors, but Tomáš does not deliver a clear set of dimensions against which related work and the newly proposed approach could be compared to.

Section 3 determines the core of the thesis, especially the solutions proposed to answer the research challenges as posed above. Unfortunately, Tomáš’ responses do focus initially on the NEMEA framework, which was introduced in earlier chapters of this thesis. Thus, the scientifically relevant focus of a thesis remains a bit hidden within Section 3. The content of NEMEA — once called framework and once termed system or even tool —, however, is the key of this work and develops well beyond state-of-the-art knowledge, which was researched, obtained, designed, and evaluated. Thus, the flow-based analysis is explained by its essential features (application-awareness and stream-wise processing). Key terminology important for this work is defined, especially “witness” (an abstract representation of semantic relations in the flow data, significant in parallel processing). And the application of these terms onto the detection algorithms gains the major observations from this thesis, well beyond details, which are not part of the referenced set of eight papers published earlier. The set of papers written by Tomáš are integrated into the flow of arguments, such that the major and minor elements of the new approach are clearly interrelated.

Furthermore, in Section 4 Tomáš summarizes and partially concludes his work, and proposes next steps. While more general summarizing elements are discussed on a per-challenge basis, the real conclusions of his work are not made that explicit. The advantages reached by the work performed are not made explicit within the thesis, but within the papers, thus, conclusions in terms of user, provider, or operator views are a bit hidden. The key observation, however, is made very clear: “This research showed that some advanced security threats are invisible for the traditional NetFlow tools.” And the solution provided within the thesis perfectly well demonstrates that these application-aware detection methods studied and designed are capable to overcome this lack of visibility! And very good for a systems-based thesis is the deployment success of the NEMEA within the Czech National Research and Education Network (NREN) — CESNET2. Unfortunately, the view into the future is very short and only states three out of the many interesting steps possible, without any minimal argumentation, why these three steps may add further advantages to the solutions and research findings already achieved.

Finally, the appendix includes the set of eight papers, which outline, define, and refine a number of aspects, relevant for a suitable answer toward the thesis’ goals. Thus, the cumulative type of a thesis submission is well documented by these papers attached.

**Evaluation and Specific Comments**

In summary, Tomáš’ work is very timely and was driven by a dedicated problem (theoretically and practically motivated at the same time) in which the technology of flow-based analysis in a real-time manner, while applying a stream-wise approach in terms of parallel analysis steps, was successful. These features are used in the detection modules designed and developed to evaluate the concept on data, which is originates from real network data.

It is a key strength of Tomáš’ work that the core of his work is a combination of formalized detection algorithms and metrics (semantic relations termed “witnesses”, and parallel flow data processing based on a Flow Scatter component) and its evaluation as well as its application in an operational prototype, called tool, too.

On one hand, Tomáš' thesis is broad in scope (flow-based analysis applied to anomaly detection), but it is well constrained (by the stream-wise view being prototyped in a parallel compute model), too. Besides the clear scientific perspective (not perfectly described in the thesis, well done within the papers, but excellently evaluated in the thesis) of this thesis' advances are very well documented beyond the state-of-the-art, especially the work in the application-aware stream-wise flow data processing. This needs to be considered to handle the practically complex approach, but being operationally useful. On the other hand, the prototype designed and evaluated is operational and well set-up and serves for way more than just a pure proof-of-concept, since it was experimented on real-life data and within CESNET2.

Tomáš' combination of methods chosen, experiments/evaluations and analysis performed, and interpretations being undertaken is well documented. His approach in this thesis is fully in-line with instruments and standards of the scientific communities of networking, which is well documented by the fact, too, that Tomáš published his research results in 5 first and 3 second author papers in peer-reviewed conferences and workshops (AIMS 2017, CNSM 2016, AIMS 2016, AIMS 2015, and IEEE CAMAD 2014). Overall, this is an excellent result for the duration of the thesis, since 9 additional conference papers have been published besides the thesis' topic.

**Recommendation**

Overall, this thesis contributed in detail to those areas listed above, for which these findings are new, unique, and evaluated in many details throughout Tomáš Cejka's written thesis. The excellent combination of theoretical and practically applicable results, combined with a thorough and in-depth evaluation — again both based in real-life data and with the running/operational system experiments — leads to a highly relevant, applicable, and very well founded knowledge advance well beyond today's state-of-the-art. The written thesis may contain, mainly due to the use of those papers considered relevant, a few inconsistencies with respect to terminology applied.

Based on this evaluation of the Ph.D. thesis submitted, Tomáš's very good performance in the thesis, and his descriptive, extensive, and constructive work, I can state that at all in all, the work determines a very good Ph.D. thesis.

Tomáš Cejka, the author of the thesis, proved the ability to conduct research and achieved scientifically valid results. Thus, I do recommend to the Czech Technical University in Prague this dissertation for the presentation and defense with the aim of receiving a Ph.D. degree. While evaluating the overall work seen, a very good grade should be considered to be awarded.

Kind regards,

Prof. Dr. Burkhard Stiller