CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Information Technology

Proceedings of the 6th Prague Embedded Systems Workshop

June 28-30, 2018

Roztoky u Prahy, Czech Republic

ISBN 978-80-01-06456-6

Editors:

doc. Ing. Hana Kubátová, CSc.

doc. Ing. Petr Fišer, Ph.D.

Ing. Jaroslav Borecký, Ph.D.

Message from the Program Chairs

The Prague Embedded Systems Workshop is a research meeting intended for the presentation and discussion of students' results and progress in all aspects of embedded systems design, testing, and applications. It is organized by members of the Department of Digital Design at Faculty of Information Technology (which is the youngest one) of the Czech Technical University in Prague (which is the oldest technical university in Central Europe). The workshop aims to enhance collaboration between different universities not only inside EU. It will be based on oral presentations, mutual communication, and discussions.

Modern embedded devices are equipped with communication interfaces and the importance of secured communication grows. Devices are usually connected into a computer network so they become a part of the network infrastructure. Besides benefits and advantages, there are security aspects and vulnerabilities that must be covered. Therefore, there is an intersection of design and development of embedded devices and an area of network security. To follow the current trends and to focus discussions on security topics, PESW organizes a special session on Network security. This section is organized by Tomáš Čejka.

There were 21 papers submitted this year, three were full papers and 18 abstracts of long presentations. Therefore, more emphasis was put on presentation, rather than publication, which is the intent of PESW. Papers from Czech Republic, Israel and Italy were present this year.

The technical program is also highlighted by four keynote speakers in the areas of testing, reliability, and cybersecurity:

- Adaptive Test Cost and Quality Optimization. Speaker: Alex Orailoglu
- Cross-Layer System-Level Reliability Estimation. Speaker: Alberto Bosio
- Increasing system reliability for safety-critical applications. Speaker: Ernesto Sanchez
- Cisco is no longer just a networking company. Speaker: Milan Habrcetl

The 6th PESW includes a student competition of the best master and bachelor diploma theses of projects close to the embedded systems area. It is organized by the IEEE Student Branch at CTU. PESW program committee members will evaluate the posters and their oral presentations to select the best ones. The winner prizes are sponsored by IEEE, STMicroelectronics, ASICentrum, CZ.NIC, ESET and CESNET.

Six technical sessions were formed, with the following topics:

- Fault tolerance (3 papers)
- Testing (3 papers)
- Signal processing (2 papers)
- Embedded systems, Emerging technologies, Modeling (3 papers)
- Communication Networks and IoT (3 papers)
- Stream-Wise Detection and Mitigation (4 papers)
- Trust and Reputation (3 papers)

Last but not least we would like to thank to our sponsors (CTU in Prague, EaToN company, ASICentrum, STMicroelectronics, CZ.NIC, ESET, CESNET and Czechoslovakia Section of IEEE).

We wish you to spend fruitful and communicative time in Roztoky.

Hana Kubátová and Petr Fišer

Committees

Workshop Chairs

Hana Kubátová, CTU in Prague (CZ)

Petr Fišer, CTU in Prague (CZ)

Programme Committee

P. Bernardi, Politecnico di Torino (IT)

A. Bosio, LIRMM, Montpellier (FR)

T. Čejka, CTU in Prague (CZ)

G. Di Natale, LIRMM, Montpellier (FR)

P. Fišer, CTU in Prague (CZ)

K. Jelemenská, STU Bratislava (SK)

P. Kitsos, TEI of Western Greece (GR)

H. Kubátová, CTU in Prague (CZ)

I. Levin, Tel-Aviv University (Israel)

A. McEwan, University of Leicester (UK)

M. Novotný, CTU in Prague (CZ)

S. Racek, UWB, Pilsen, (CZ)

E. Sanchez, Politecnico di Torino (IT)

J. Schmidt, CTU in Prague (CZ)

M. Skrbek, CTU in Prague (CZ)

R. Stojanovic, Univ. of Podgorica (ME)

J. Strnadel, BUT, Brno (CZ)

R. Ubar, Tallinn Univ. of Technology (EE)

H. T. Vierhaus, BTU Cottbus (Germany)

Special Session on Network Security Chair

Tomáš Čejka, CTU in Prague (CZ)

Student Poster Session Chair

Jan Bělohoubek, CTU in Prague (CZ)

Organizing Committee

H. Kubátová, CTU in Prague (CZ)

P. Fišer, CTU in Prague (CZ)

R. Kinc, AMCA (CZ)

E. Uhrová, AMCA (CZ)

Contents

Keynotes

Increasing system reliability for safety-critical applications

Speaker: **Ernesto Sanchez**, *Politecnico di Torino, Italy*

Today, safety- and mission-critical applications are asking for increasing the system dependability during the operational lifetime. Actually, new standards arose in the last years try to define the minimum requests in order to guarantee reliability of such devices. In fact, during the last years, microprocessor-based safety-critical applications are introducing a series of audit processes to be applied during the whole product lifetime targeting reliability. Some of these processes are common in industrial design and manufacturing flows, including risk analysis, design verification, and validation, performed since the early phases of product development, but very often, additional test processes need to be performed during the product mission life in a periodic fashion to match reliability standards. In this talk, a brief guideline to effectively increase system dependability by exploiting functional approaches is provided. The most important constraints that need to be considered during the generation phase, as well as during the execution time are described. Additionally, a comparison checking three different strategies on a particular module of an industrial pipelined processor core is also provided.

Ernesto Sanchez

Ernesto Sanchez received his degree in Electronic Engineering from Universidad Javeriana - Bogota, Colombia in 2000. In 2006 he received his Ph.D. degree in Computer Engineering from the Politecnico di Torino, where currently, he is an Associate Professor with Dipartimento di Automatica e Informatica. His main research interests include evolutionary computation, functional microprocessor verification, validation, and testing.

Cisco is no longer just a networking company

Speaker: **Milan Habrcetl**, *Cisco CyberSecurity Specialist, Praha, Czech Rep.*

Cybersecurity has become the phenomenon of the present era. Without data protection and infrastructure, it's hard to achieve of a prosperous business. Let's have a look how cybersecurity can be more efficient and automated, and what challenges await us in the near future.

Milan Habrcetl

Since 1998, he has been deeply involved in cyber security industry, mostly in positions of sales manager or business development manager. He has been working in Cisco since February 2016 within the Global Security Sales organization to support the sales of entire Cisco security portfolio in the Czech Republic and Slovakia.

Cross-Layer System-Level Reliability Estimation

Speaker: **Alberto Bosio**, *LIRMM Montpellier, France*

Cross-layer approach is becoming the preferred solution when reliability is a concern in the design of a microprocessor-based system. Nevertheless, deciding how to distribute the error management across

the different layers of the system is a very complex task that requires the support of dedicated frameworks for cross-layer reliability analysis. In other words, the designer has to know what are the "critical" components of the system in order to properly introduce error management mechanisms. Unfortunately, system-level reliability estimation is a complex task that usually requires huge simulation campaign. This presentation aims at proposing a cross-layer system-level reliability analysis framework for soft-errors in microprocessor-based systems. The framework exploits a multi-level hybrid Bayesian model to describe the target system and takes advantage of Bayesian inference to estimate different reliability metrics.

Experimental results, carried out on different microprocessor architectures (i.e., Intel x86, ARM Cortex-A15, ARM Cortex-A9), show that the simulation time is significantly lower than state-of-the-art fault-injection experiments with an accuracy high enough to take effective design decision.

Alberto Bosio

Alberto Bosio received the PhD in Computer Engineering from Politecnico di Torino in Italy in 2006 and the HDR (Habilitation Diriger les Recherches) in 2015 from the University of Montpellier (France). Currently he is an associate professor in the Laboratory of Informatics, Robotics and Microelectronics of Montpellier (LIRMM)-University of Montpellier in France. He has published articles in publications spanning diverse disciplines, including memory testing, fault tolerance, diagnosis and functional verification. He is an IEEE member and the chair of the European Test Technology Technical Council (ETTTC).

Adaptive Test Cost and Quality Optimization

Speaker: **Alex Orailoglu**, *University of California, San Diego, USA*

The higher levels of integration and process scaling impose failure behaviors which are challenging to interpret, necessitating the continuous augmentation of fault models and test vectors in the hopes of taming the defect escape rate. The subsequent inflation in the number of test vectors coupled with the constant increase in the size of each test vector continuously boosts test cost. The economics of particularly the competitive consumer marketplace however require a constant vigilance at the test cost while ensuring a satisfactory test quality.

While the inclusion of new fault models helps boost test quality, the non-uniform distribution of various defect types and the defect coverage overlaps between fault models imply variable effectiveness of fault models and test vectors, resulting in the inclusion of a large number of ineffective vectors in test flow. A static derivation of test effectiveness however remains problematic in practice as it is well known that defect characteristics are prone to drifts throughout the product lifecycle. Furthermore, the increasing process variation and the integration of hundreds of domains within a chip result in increasingly distinct domains and individualized chip instances with diverse test resource requirements. The conventional test method of a static application of an identical test set to all chips consequently struggles to satisfy the demanding test cost and quality constraints in the face of the evolving defect behaviors and the increasing diversification in test resource requirements.

This talk addresses the simultaneous necessity for satisfactory test quality and low test cost through an adaptive test cost and quality optimization framework. The proposed methodologies not only adaptively assess the effectiveness of fault models and test vectors but also evaluate the variable test resource requirements of the chips and domains based on their distinct characteristics, enabling an effective yet efficient test through the selection of the most effective vectors and a carefully crafted allocation of test resources. The proposed methodologies are tailored for a broad set of application scenarios through the consideration of different defect classes and defect characteristic drift types while incorporating the test data gathering

and delivery constraints and overcoming the associated algorithmic challenges.

Alex Orailoglu

Alex Orailoglu received his S.B. Degree cum laude in applied mathematics from Harvard College, Cambridge, MA, and the M.S. and Ph.D. degrees in computer science from the University of Illinois at Urbana-Champaign, Urbana.

He is currently a Professor of Computer Science and Engineering with the Department of Computer Science and Engineering, University of California, San Diego, where he directs the Architecture, Reliability and Test (ART) Laboratory, focusing on VLSI test, computer architectures, reliability, embedded processors and systems, and nanoarchitectures. He has published more than 250 papers in these areas.

Dr. Orailoglu has served as the General Chair and the Program Chair for the IEEE/ACM/IFIP International Symposium on Hardware/Software Codesign and System Synthesis, the IEEE VLSI Test Symposium, the IEEE Symposium on Application-Specific Processors (SASP), the Symposium on Integrated Circuits and Systems Design (SBCCI), the IEEE/ACM International Symposium on Nanoscale Architectures (NanoArch), the HiPEAC Workshop on Design for Reliability and the IEEE International High Level Design Validation and Test Workshop (HLDVT). He has last served as the Program Co-Chair of IFIP/IEEE International Conference on Very Large Scale Integration (VLSISoC) 2013. He has co-founded the IEEE SASP, the IEEE/ACM NanoArch, the IEEE HLDVT, and the HiPEAC Workshop on Design for Reliability.

Professor Orailoglu has served as a member of the IEEE Test Technology Technical Council (TTTC) Executive Committee, as the Vice Chair of TTTC, as the Chair of the Test Technology Education Program group, as the Technical Activities Committee Chair and the Planning Co-Chair of TTTC and as the Communities Chair of the IEEE Computer Society Technical Activities Board. He is the founding chair of the IEEE Computer Society Task Force on Hardware/ Software Codesign and the founding vice-chair of the IEEE Computer Society Technical Committee on NanoArchitectures.

Dr. Orailoglu has served as an IEEE Computer Society Distinguished Lecturer. He is a Golden Core Member of the IEEE Computer Society.

Problems of a Software Test Library for Multicore System-On-Chip

Paolo Bernardi, Andrea Floridia, Davide Piumatti, Ernesto Sanchez
Politecnico di Torino – Dipartimento di Automatica e Informatica
Corso Duca degli Abruzzi 24, Torino (Italy)

paolo.bernardi@polito.it
andrea.floridia@polito.it
davide.piumatti@polito.it
ernesto.sanchez@polito.it

Keywords. Parallel test. In-field testing. Multi-core Software-Based Self-Test.

Abstract

In recent years the complexity of System-On-Chips growth exponentially, mainly due to the ever-increasing demand for more functionalities, even for embedded applications. In order to fulfil such requests, semiconductor vendors introduced in this market multi-core devices to satisfy the more complex software algorithms use to image recognition for implementing the Advanced Driver Assistance Systems. However, despite the gain in terms of performance, the adoption of multi-core devices poses several issues from the test point of view. In particular, it is necessary to evolve the in-field test strategies (commonly used to increase the reliability level of a processor-based system) from the single core to the multi-core case. We present a possible approach for rapidly migrating a Software Test Library (STL), developed according the Software-Based Self-Test (SBST) approach for a single-core processor, to a multi-core processor. The solution proposed use the hardware semaphores in order to control the access to shared resources among different cores. This approach requiring a minimal modification of the test programs, yet without affecting the fault coverage detected by the STL. The hardware semaphores were exploited in order to implement the parallel execution of the programs among different cores, a precalculated scheduler order is need to optimize the total execution time of all STL on all core of the microcontroller.

1.1 Multi core programming and Scheduling problems for the STL

In literature are present different approach to executing a STL in parallel on different cores integrated in the same chip. Initially, it is necessary to distinguish among the multiprocessor chip (CMP) [1] and multithread chip (CMT) [3] architecture. The first type is composed of N replicas of the same core. In the CMP chip the different cores communicated through a shared memories hierarchy, in [2] a possible technique of parallel test suite is proposed. The solution is based on a shared two-level cache memory hierarchy, al L1 private and a L2 shared, for executing the STL and for reduce the access to the flash memory. A scheduling algorithm is also proposed for reducing the execution time. In the CMT architecture, a particular mechanism allows a fast contest switch between two threads in execution. This is possible because a few hardware units inside of the core are duplicated (as the register file or the cache memory) and a selector enable the copy associated to the thread currently in execution. In [4] a method for splitting test routines among the available threads is proposed, aiming at reducing the core idle intervals. In [5] a methodology targeting both optimization of test execution time and improvement of the fault coverage is described. In the case in question uses a CMP chip with the use of hardware semaphores to synchronize the cores.

Multi-core systems often include some hardware resources shared among different cores as memories or peripherals. It is necessary a safe mechanism for controlling the access to such resources. Possible mechanism are hardware semaphores that implement a simple mode to "lock and unlock" the shared hardware resources. The same STL software, in parallel execution on different cores, uses the semaphores for acquire temporary exclusive use. The others core remains in wait state until the resource is not released. In general, a hardware semaphore must guarantee the exclusive use to the resource only to the processor that has executed the "lock" operation and forbid the "lock" operation to the others process. Moreover, the processor locking a particular shared resource is the only one that can unlock it. From the programmer's perspective, hardware semaphores are seen just like a peripheral. The basic operations that can be performed on a semaphore are the CheckStatusLock, the LockSem and the UnlockSem. The CheckStatusLock is useful to checks whether a particular semaphore is locked or not, while the LockSem and UnlockSem are useful for implementer the "lock and unlock" mechanism previous described.

Traditional scheduling algorithms are not applicable in the STL case, since all tests must be performed on all system cores. In general, traditional planning considers the activity performed when it is entirely performed on only one core. This constraint increases the complexity of the problem. In addition, the problem of shared resources must be considered to avoid waiting states between two or more processes. In general, the STL is executed at system startup, first on the execution of the operating system or customer code. In boot-time no other codes are executed and a precalculated optimized test order are scheduling. Aim of scheduling is to reduce the total execution time on all the cores.

1.2 Conclusion

We proposed approach for migrating a STL, originally developed for only one of the cores composing a multi-core SoC, to the whole set of cores. The proposed methodology requires only the availability of hardware semaphores in order to enable the concurrent execution of test programs among different cores. The approach does not impact to the fault coverage and required to easy modify to the STL code for use the semaphores. Clearly, the higher the number of test programs composing the STL and the higher the number of conflicts among them for use the shared resources. In this situation is harder is to find the optimal solution that reduces the overall test execution time.

Paper origin

Parallel Software-Based Self-Test suite for Multi-core System-on-Chip: migration from single-core to multi-core automotive microcontrollers. A. Floridia, D. Piumatti, E. Sanchez, S. De Luca, A. Sansonetti This paper has been accepted and presented at the 13th IEEE International Conference on" Design & Technology of Integrated Systems in Nanoscale Era" (DTIS) 2018 April 10-12, 2018, Taormina, Italy.

References

[1] Hammond, Nayfeh, Olukotun, "A single-chip multiprocessor", Computer, Vol. 30, Issue 9, Sep. 1997, pp 79 – 85

[2] Apostolakis, Gizopoulos, Psarakis, Paschalis, "Software-Based Self-Testing of Symmetric Shared-Memory Multiprocessors", IEEE Transactions on Computers, Dec. 2009, Vol. 58, Issue 12, pp. 1682 - 1694

[3] Chaudhry, Cypher, Ekman, Karlsson, Landin, Yip, Tremblay, "Rolc: A High-Performance Sparc CMT Processor", IEEE Micro, Apr. 2009, Vol. 29, Issue 2

[4] Apostolakis, Psarakis, Gizopoulos, Paschalis, Parulkar, "Exploiting Thread-Level Parallelism in Functional Self-Testing of CMT Processors", Test Symposium, 2009 14th IEEE European, 25-29 May 2009

[5] Foutris, Psarakis, Gizopoulos, Apostolakis, Vera, Gonzalez, "MT-SBST: Self-Test Optimization in Multithreaded Multicore Architectures", IEEE International Test Conference (ITC), 2-4 Nov. 2010, Austin, TX (USA)

Development flow of on-line Software Test Libraries for asynchronous processor cores

Andrea Floridia, Ernesto Sanchez
Politecnico di Torino
Torino, Italy

andrea.floridia@polito.it,
ernesto.sanchez@polito.it

Keywords. Software-Based Self-Test, on-line testing, desynchronization.

Abstract

Asynchronous design style is quite appealing from different perspectives. Several studies confirmed the reliability of asynchronous circuits in harsh environments, being capable to better tolerate power supply and temperature variations with respect to the synchronous counterparts. However, despite these advantages and many others, their applicability (especially in safety-critical scenarios) is today quite limited. In addition, commercial EDA tools can be hardly applied to most of the asynchronous designs; therefore, designers are discouraged to use such devices in their applications. Notably, devices deployed for safety-critical applications must satisfy stringent requirements to guarantee the highest reliability level, defined for example in the ISO 26262 standard for automotive applications. Commonly, on-line testing mechanisms are necessary to achieve such requirements (e.g., On-line Built-in Self-Test, Lockstep Execution, and Software Test Libraries). Such mechanisms undergo several validation processes to assess their effectiveness, being fault injection campaigns the most commonly used. It is important to note that for such procedures, designers exploit standardized commercial EDA tools, intended to certificate standards compliance. In this study, we describe a methodology for the development and the evaluation of Software Test Libraries (STLs) targeting the on-line testing of asynchronous processor cores, using exclusively commercial tools used by industries for the functional safety analysis. Currently, we are targeting stuck-at faults, but the proposed flow can be extended to other fault models. The proposed flow starts from a synchronous processor core; then, a preliminary step consisting in the desynchronization [2] of the processor is performed. The selected design methodology is fully compatible with EDA tools and it does not require a detailed knowledge of asynchronous design. Concerning the STL, these programs are developed according to the Software-Based Self-Test (SBST) approach. Our case study is the DLX processor used in the ASPIDA [1] project. In the literature, there exist many SBST strategies targeting synchronous-based processor cores, along with well-established methodologies for assessing their effectiveness. For synchronous processors, the fault simulator, for example, is instructed to periodically observe some meaningful system bus signals at specific time instants (hereinafter strobe points). In those time instants, the processor is supposed to store the signature produced by the test program execution in the available memory. Each strobe point is characterized by a period of observation and a time offset. The period of observation, called strobe period, is equivalent to the processor clock period, while the time instant in which values should be stable and the signature is being stored into the data memory is called strobe offset. Usually, the strobe offset is computed so that the observation of such signals is performed very close to the end of the clock period, a time instant in which the circuit is supposed to be in a quiescent state (i.e., not switching anymore). It is worth noting that strobe points are not necessary contiguous time instants (i.e., one immediately after the previous). In fact, to reproduce the same scenario of the on-line test, the test engineer should compute strobe offset and period for each strobe point so that system bus signals are

observed exclusively when they hold the signature (otherwise, too optimistic fault coverages could be obtained). Intuitively, the lack of a clock signal makes the application of such a strategy to asynchronous processors not trivial as is. However, the handshake controllers of the desynchronized version behave as local clock generators for each stage of the processor pipeline. These clocks are the enable signals for the sequential elements (namely the latches) of the stage they are associated with. Such signals are continuously generated by the controllers, according to a given handshake protocol. Thus, by carefully observing the behavior of the desynchronized processor, we could conclude that the memory stage of the processor is periodically enabled. Indeed, even though the memory access is not actually performed (no valid data present on the bus), the controller generates the enable signals for that stage. Since this stage interfaces the processor with the data memory, is the only portion of the circuit through which the test signature can be observed (namely, when it is stored in memory). The data are stored in the slave latch when the enable signal goes from high to low (mimicking the edge-triggered behavior of a synchronous circuit). Bus data lines are assumed to be stable before slave latch closes. Hence, by synchronizing the fault simulator with that signal period, it is possible to use the very same methodology for the asynchronous processors as well. In the following, the proposed fault simulation flow is described. Starting from the desynchronized processor, both the STL and the processor are simulated resorting to a commercial logic simulator. From this initial simulation, the strobe points for the fault simulator are extracted. It is important to note that in a synchronous processor, this step is not required since strobe points are synchronized with the system clock. At this point, the fault simulation campaign begins. The fault simulator receives as inputs the design, the STL, strobe points and an initial fault list. It is worth mentioning that also the design Standard Delay Format (SDF) file is needed (unlike the synchronous circuits that do not require detailed timing information for the fault simulation) to correctly simulate the processor model. Additionally, using real delays for fault simulation prevents potential issues arising from combinational feedback loops (i.e., infinite execution in zero simulation time). After the first run, the outcome of the fault simulator is analyzed to identify possible discarded faults. Among these faults, it is likely to find *hyperactive* faults (*HA* faults). When dealing with asynchronous processors these faults are more relevant due to the asynchronous control network. Moreover, to obtain the compliance with safety standards, all the faults must be categorized as detected or not detected. Thus, a second run is needed during which the fault list is updated with the *HA* faults only. At the same time, the fault simulator should be configured to do not discard them during the fault simulation. The proposed methodology has been evaluated using two test programs targeting the processor adder and register file, implementing well-kwon test algorithms originally developed for synchronous processors. As functional safety fault simulator, ZO1X by Synopsys has been used. By adopting the aforementioned flow, we were able to obtain the very same fault coverage as in the synchronous case. Thus, it is possible to reuse the clear majority of the already existing algorithms for processor functional testing. However, the time required for completing the fault injection campaign is increased due to the necessity of SDF file (which slows down fault simulation). We are currently working to extend this methodology to include also delay faults, and to other processors, obtained with different design techniques.

Paper origin

This paper has been accepted and will be presented at the conference 24th IEEE International Symposium on On-line Testing and Robust System Design.

References

[1] http://www.ics.forth.gr/carv/aspida/

[2] J. Cortadella, A. Kondratyev, L. Lavagno and C. P. Sotiriou, "Desynchronization: Synthesis of Asynchronous Circuits From Synchronous Specifications," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Oct. 2006. doi: 10.1109/TCAD.2005.860958

ZATPG: SAT-based ATPG for Zero-Aliasing Compaction

Robert Hülle, Petr Fišer, Jan Schmidt
Faculty of Information Technology, Czech Technical University in Prague
Prague, Czech Republic

{hullerob,fiserp,schmidt}@fit.cvut.cz

Keywords. ATPG, aliasing, compaction, LFSR, MISR, SAT, zero aliasing

Abstract

One of long-standing problems in digital circuit testing is fault aliasing in the response compaction. Fault aliasing is an important source of coverage loss, especially if we strive to achieve high compaction ratio. Existing methods to lower or eliminate aliasing mostly require changes to the compactor design. This can lead to a higher compactor complexity, bigger area overhead, longer propagation paths, etc.

We propose a method to eliminate aliasing without the need to modify the compactor design. The basic idea is to constrain the test pattern generation itself to produce a test with zero aliasing. This is in contrast to previous methods, where a test is computed independently and the anti-aliasing algorithm does not modify the test further [1, 2]. Some anti-aliasing algorithms exert a partial control over a test sequence, by reordering (already existing) test [3–5].

Note that we are only considering aliasing in a temporal compactor. Preventing aliasing in a spatial compactor is much easier problem, for both pre-existing and new test set. In our paper, we assume a spatial compactor that does not introduce new redundant faults.

1.1 Constraining the ATPG

Our method, ZATPG (zero-aliasing test patterns generator), is based on a SAT-based (Boolean satisfiability) ATPG (automated test patterns generator) [6]. Conventional SAT-ATPG works by modelling a fault as a replica of the CUT, transforming the miter to a CNF (conjunctive normal form), and solving the resulting CNF-SAT problem with a SAT solver [7].

We then expand the miter by introducing anti-aliasing constraints in the following way. First, we construct the miter as usual, consisting of the fault-free circuit and a circuit with the tested fault f_i and find a test pattern p_i to detect f_i. Additionally, we insert selected faults $f_{s,1}$–$f_{s,m}$, modelled in their own replicas of the CUT (Figure 1).

Aliasing happens only *after* the application of test pattern that causes it. It is therefore necessary to know the future state of compactor during the generation of test pattern p_i. This is achieved by unrolling combinational part of the compactor (block *MISR* in Figure 1). Previous state of the compactor also needs to be supplied ($S_f f$,S_1–S_m) and the output (next state, partial signature) is constrained to differ from the state for the fault-free circuit.

1.2 Results

For experiments, we have used a slight simplification possible for linear compactors (for details see [8]). The experiments were performed on benchmark circuits from the ISCAS'85 and selected ITC'99

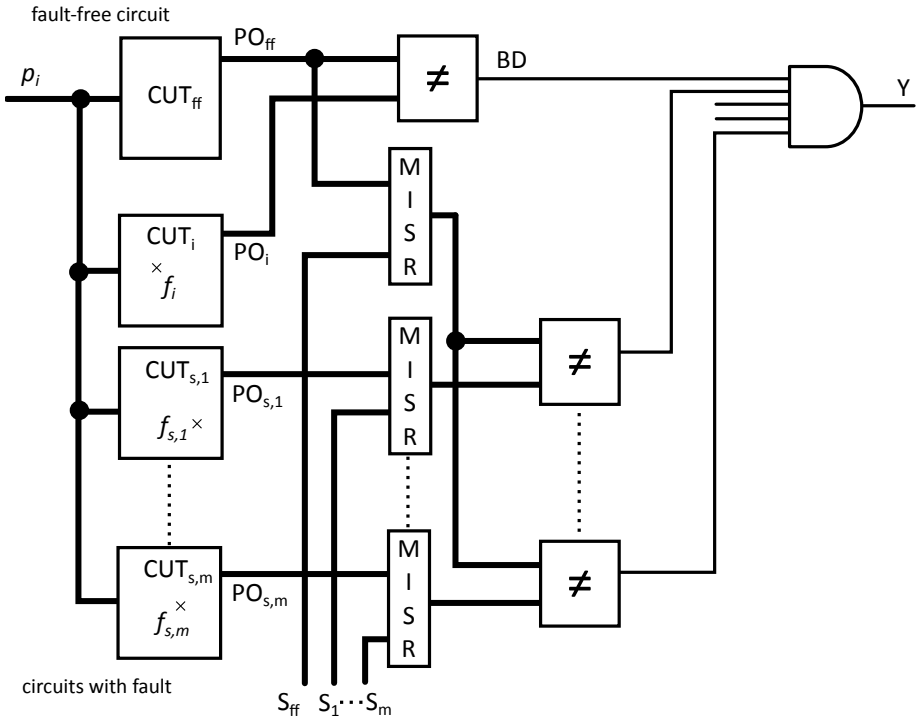

Figure 1: Extended miter for zero-aliasing in compaction

benchmarks.

For all tested circuits, we have achieved zero aliasing (full coverage) with ZATPG for smaller compactors of the same design (LFSR) than with normal ATPG. The observed gain was between 2 and 5 bits of LFSR saved. With the exception of the circuit c7552, where the achieved LFSR size was same as with ATPG (Table 1).

Paper origin

This paper has been accepted and published in the *Microprocessors and Microsystems*, vol. 61, 2018 [8].

Acknowledgment

This work was partially supported by the grant GA16-05179S of the Czech Grant Agency, "Fault Tolerant and Attack-Resistant Architectures Based on Programmable Devices: Research of Interplay and Common Features" (2016-2018).

The authors acknowledge the support of the OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 "Research Center for Informatics"

This research has been in part supported by CTU grant SGS17/213/OHK3/3T/18.

Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures"

Table 1: Minimal size of MISR with zero-aliasing test

		LFSR-MISR		CA-MISR 90/150	
circuit	faults	ATPG	ZATPG	ATPG	ZATPG
b04	2846	10	7	10	8
b11	2382	11	6	10	9
c499	970	15	8	10	10
c880	1582	12	5	10	6
c1355	2618	11	8	12	8
c1908	2581	10	8	10	8
c2670	3613	11	7	12	6
c5315	7964	12	7	12	10
c7552	10921	8	8	9	10

References

[1] K. Pradhan, D. and K. Gupta, Sandeep, "A new framework for designing and analyzing BIST techniques and zero aliasing compression," *IEEE Transactions on Computers*, vol. 40, no. 6, pp. 743–763, 1991.

[2] M. Kopec, "Can nonlinear compactors be better than linear ones?" *IEEE Transactions on Computers*, vol. 44, no. 11, pp. 1275–1282, Nov. 1995.

[3] G. Edirisooriya and P. Robinson, John, "Test generation to minimize error masking," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, no. 4, pp. 540–549, April 1993.

[4] G. Edirisooriya, P. Robinson, John, and S. Edirisooriya, "On the performance of augmented signature testing," in *IEEE International Symposium on Circuits and Systems*, May 1993, pp. 1607–1610.

[5] T. Bogue, M. Gossel, H. Jurgensen, and Y. Zorian, "Built-in self-test with an alternating output," in *Proceedings Design, Automation and Test in Europe*, Feb. 1998, pp. 180–184.

[6] R. Hülle, P. Fišer, J. Schmidt, and J. Borecký, "SAT-ATPG for application-oriented FPGA testing," in *15th Biennial Baltic Electronics Conference*, Oct. 2016, pp. 83–86.

[7] T. Larrabee, "Test pattern generation using Boolean satisfiability," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 11, no. 1, pp. 4–15, Jan. 1992.

[8] R. Hülle, P. Fišer, and J. Schmidt, "ZATPG: SAT-based test patterns generator with zero-aliasing in temporal compaction," *Microprocessors and Microsystems*, vol. 61, pp. 43 – 57, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0141933118300966

A HYBRID DSP/DEEP LEARNING APPROACH TO REAL-TIME FULL-BAND SPEECH ENHANCEMENT

Sagy Harpaz
UVEYE ltd.
Menachem Begin rd. Tel Aviv
6522042, Israel
sagy.harpaz@gmail.com

Avihay Eini
Afeka Tel Aviv Academic
College of Engineering
Bney Efraim rd. Tel Aviv
69107, Israel
avihayeini@gmail.com

Gabi Shafat
Afeka Tel Aviv Academic
College of Engineering
Bney Efraim rd. Tel Aviv
69107, Israel
gabis@afeka.ac.il

Abstract. Despite noise suppression being a mature area in signal processing, it remains highly dependent on fine tuning of estimator algorithms and parameters. In this paper, we demonstrate a hybrid DSP/deep learning approach to noise suppression. A deep neural network with four hidden layers is used to estimate ideal critical-band gains, while a more traditional pitch filter attenuates noise between pitch harmonics. The approach achieves significantly higher quality than a traditional minimum mean squared error spectral estimator, while keeping the complexity low enough for real-time operation at 48 kHz on a low-power processor.

Index Terms— noise suppression, deep learning

1 Introduction

Noise suppression has been a topic of interest since at least the 70s. Despite significant improvements in quality, the high-level structure has remained mostly the same. Some form of spectral estimation technique relies on a noise spectral estimator, itself driven by a voice activity detector (VAD) or similar algorithm, as shown in Fig. 1. Each of the 3 components requires accurate estimators and are difficult to tune. For example, the crude initial noise estimators and the spectral estimators based on spectral subtraction [1] have been replaced by more accurate noise estimators [2] and spectral amplitude estimators [4]. Despite the improvements, these estimators have remained difficult to design and have required significant manual tuning effort. That is why recent advances in deep learning techniques are appealing for noise suppression. Deep learning techniques are already being used for noise suppression [5]. Most of the proposed approaches target automatic speech recognition (ASR) applications, where latency and computational resources are not important factors. The approach contrasts with so-called end-to-end systems where most or all of the signal processing operations are replaced by machine learning. These end-to-end system have clearly demonstrated the capabilities of deep learning, but they often come at the cost of significantly increased complexity.

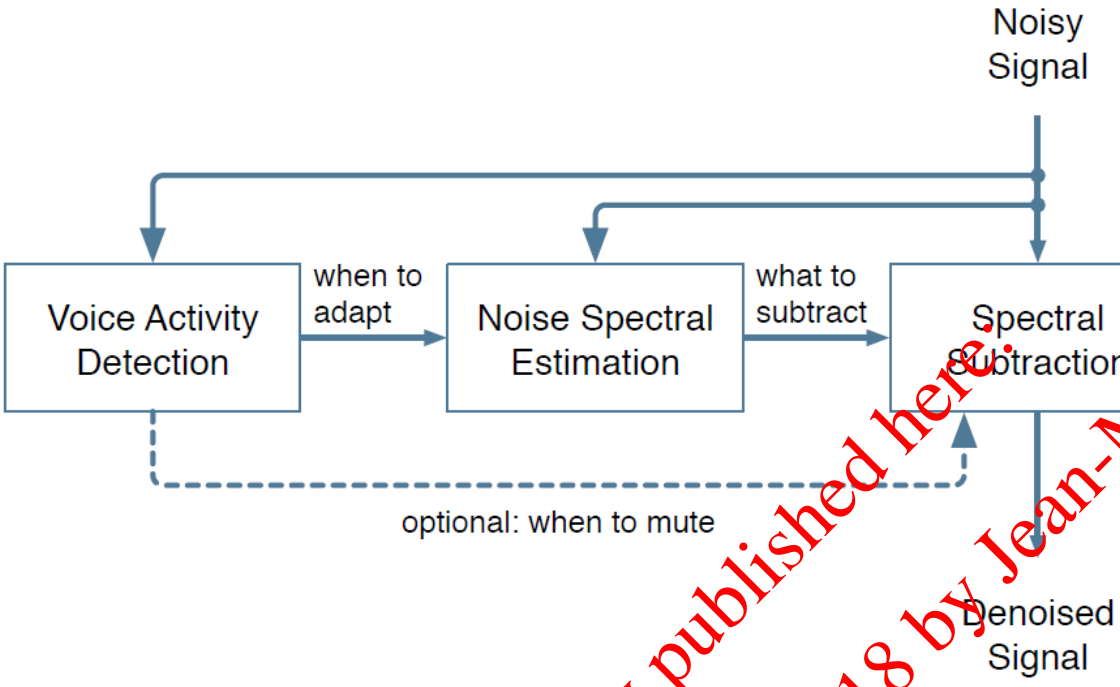

Fig. 1. High-level structure of most noise suppression algorithms.

2 Signal Model

In the proposed approach we instead focus on real-time applications (e.g. video-conference) with low complexity. We also focus on fullband (48 kHz) speech. To achieve these goals we choose a hybrid approach. The goal is to use deep learning for the aspects of noise suppression that require careful tuning while using basic signal processing building blocks for parts that do not. The main processing loop is based on 20 ms windows with 50% overlap (10 ms offset). Both analysis and synthesis use the same Vorbis window [9], which satisfies the Princen-Bradley criterion [10]. The window is defined as

$$w(n) = \sin\left[\frac{\pi}{2}\sin^2\left(\frac{\pi n}{N}\right)\right] \quad (1)$$

where N is the window length. The signal-level block diagram for the system is shown in Fig. 2. The bulk of the suppression is performed on a low-resolution spectral envelope using gains computed from a recurrent neural network (RNN). Those gains are simply the square root of the ideal ratio mask (IRM). A finer suppression step attenuates the noise between pitch harmonics using a pitch comb filter.

2.1 Band structure

A neural network is used to directly estimate magnitudes of frequency bins and requires a total of 6144 hidden units and close to 10 million weights to process 8 kHz speech. Scaling to 48 kHz speech using 20-ms frames would require a network with 400 outputs (0 to 20 kHz), which

clearly results in a higher complexity than we can afford. One way to avoid the problem is to assume that the spectral envelopes of the speech and noise are sufficiently flat to use a coarser

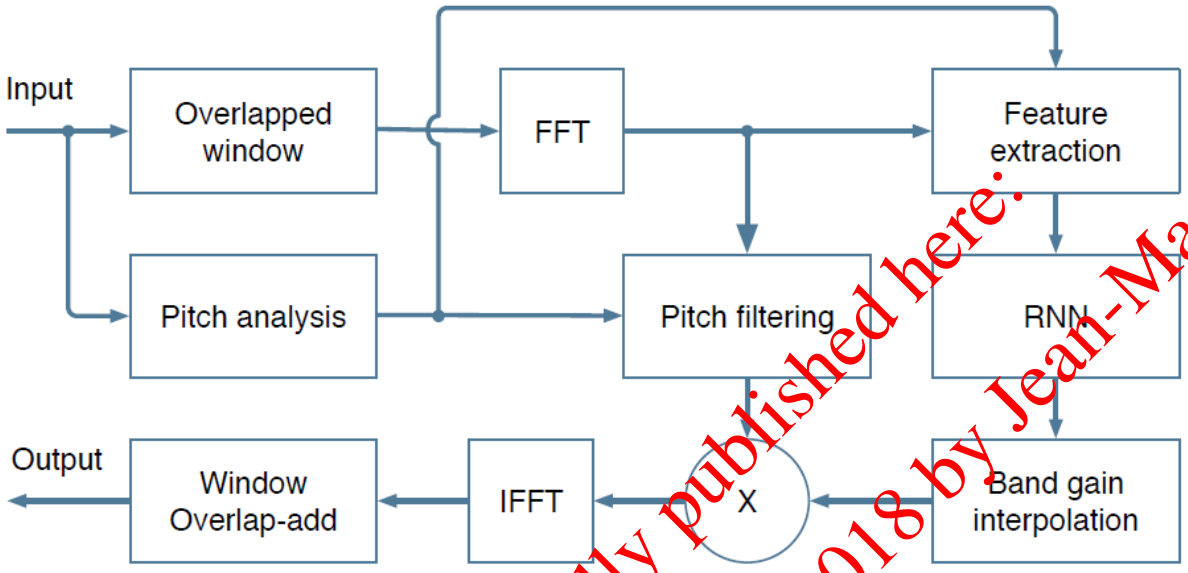

Fig. 2. Block diagram.

resolution than frequency bins. Also, rather than directly estimate spectral magnitudes, we instead estimate ideal critical band gains, which have the significant advantage of being bounded between 0 and 1. We choose to divide the spectrum into the same approximation of the Bark scale [11] as the Opus codec [12] uses. That is, the bands follow the Bark scale at high frequencies, but are always at least 4 bins at low frequencies. Rather than rectangular bands, we use triangular bands, with the peak response being at the boundary between bands. This results in a total of 22 bands. Our network therefore requires only 22 output values in the [0, 1] range. Let $\omega_b(k)$ be the amplitude of band b at frequency k, we have $\sum_b \omega_b(k) = 1$. For a transformed signal $X(k)$, the energy in a band is given by

$$E(b) = \sum_b \omega_b(k)|X(k)|^2 \tag{2}$$

The per-band gain is defined as g_b

$$g_b = \sqrt{\frac{E_s(b)}{E_x(b)}} \tag{3}$$

Where $E_s(b)$ is the energy of the clean (ground truth) speech and $E_x(b)$ is the energy of the input (noisy) speech. Considering an ideal band gain $\hat{g}_b$, the following interpolated gain is applied to each frequency bin k:

$$r(k) = \sum_b \omega_b(k)\hat{g}_b \tag{4}$$

2.2 Pitch filtering

The main disadvantage of using Bark-derived bands to compute the gain is that we cannot model finer details in the spectrum. In practice, this prevents noise suppression between pitch harmonics. As an alternative, we can use a comb filter at the pitch period to cancel the inter-harmonic noise in a similar way that speech codec post-filters operate [13]. Since the periodicity of speech signal depends heavily on frequency (especially for 48 kHz sampling rate), the pitch filter operates in the frequency domain based on a per-band filtering coefficient $\boldsymbol{\alpha_b}$. Let P(k) be the windowed DFT of the pitchdelayed signal x (n − T), the filtering is performed by computing $X(k) + \boldsymbol{\alpha_b} P(k)$ and then renormalizing the resulting signal to have the same energy in each band as the original signal X(k).
The pitch correlation for band b is defined as

$$\boldsymbol{p_b} = \frac{\sum_{\boldsymbol{k}} \omega_b(k) \Re[\boldsymbol{X}(\boldsymbol{k})\boldsymbol{P}^*(\boldsymbol{k})]}{\sqrt{\sum_{\boldsymbol{k}} \omega_b(k)|\boldsymbol{X}(\boldsymbol{k})|^2 \cdot \sum_{\boldsymbol{k}} \omega_b(k)|\boldsymbol{P}(\boldsymbol{k})|^2}} \tag{5}$$

where $\Re[\cdot]$ denotes the real part of a complex value and * denotes the complex conjugate. Note that for a single band, (5) would be equivalent to the time-domain pitch correlation. Deriving the optimal values for the filtering coefficient αb is hard and the values that minimize mean squared error are not perceptually optimal. Instead, we use a heuristic based on the following constraints and observations. Since noise causes a decrease in the pitch correlation, we do not expect $\boldsymbol{p_b}$ to be greater than $\boldsymbol{g_b}$ on average, so for any band that has $\boldsymbol{p_b} \geq \boldsymbol{g_b}$, we use $\boldsymbol{\alpha_b} = 1$. When there is no noise, we do not want to distort the signal, so when $\boldsymbol{g_b} = 1$, we use $\boldsymbol{\alpha_b} = 0$. Similarly, when $\boldsymbol{p_b} = 0$, we have no pitch to enhance, so $\boldsymbol{\alpha_b} = 0$. Using the following expression for the filtering coefficient respects all these constraints with smooth behavior between them:

$$\alpha_b = \min\left(\sqrt{\frac{p_b^2(1 - g_b^2)}{g_b^2(1 - p_b^2)}}, 1\right) \tag{6}$$

Even though we use an FIR pitch filter here, it is also possible to compute P(k) based on an IIR pitch filter of the form $H(z) = {}^{1}/_{(1-\beta_z{}^{-T})}$, , resulting in more attenuation between harmonics at the cost of slightly increased distortion.

2.3 Feature extraction

It only makes sense for the input of the network to include the log spectrum of the noisy signal based on the same bands used for the output. To improve the conditioning of the training data, we apply a DCT on the log spectrum, which results in 22 Bark-frequency cepstral coefficients (BFCC). In addition to these, we also include the temporal derivative and the second temporal derivative of the first 6 BFCCs. Since we already need to compute the pitch in (5), we compute the DCT of the pitch correlation across frequency bands and include the first 6 coefficients in our set of features. At last, we include the pitch period as well as a spectral non-stationarity metric that can help in speech detection. In total we use 42 input features. Unlike the features typically used in speech recognition, these features do not use cepstral mean normalization and do include the first cepstral coefficient. The choice is deliberate given that we have to track the absolute level of

the noise, but it does make the features sensitive to the absolute amplitude of the signal and to the channel frequency response. This is addressed in Sec. 3.1

3 DEEP LEARNING ARCHITECTUR

We rely on proven signal processing techniques and use deep learning to replace the estimators that have traditionally been hard to correctly tune. The neural network closely follows the traditional structure of noise suppression algorithms, as shown in Fig. 3. The design is based on the assumption that the three recurrent layers are each responsible for one of the basic components from Fig. 1. Of course, in practice the neural network is free to deviate from this assumption (and likely does to some extent). It includes a total of 215 units, 4 hidden layers, with the largest layer having 96 units. The increasing of the number of units does not significantly improve the quality of the noise suppression. However, the loss function and the way we construct the training data have a large impact on the final quality. We find that gated recurrent unit (GRU) [14] slightly outperform LSTM on this task, while also being simpler. Despite the fact that it is not strictly necessary, the network includes a VAD output. The extra complexity cost is very small (24 additional weights) and it improves training by ensuring that the corresponding GRU indeed learns to discriminate speech from noise.

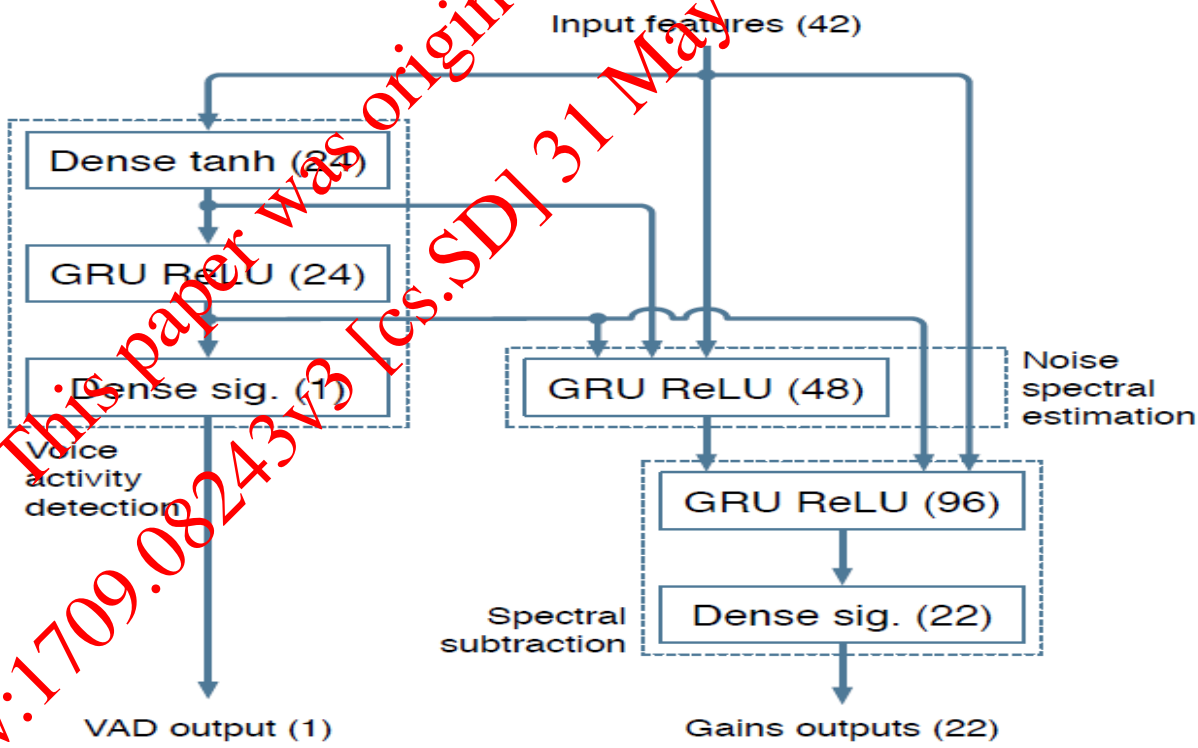

Fig. 3. Architecture of the neural network, showing the feedforward, fully connected (dense) layers and the recurrent layers, along with the activation function and the number of units for each layer.

3.1 Training data

Since the ground truth for the gains requires both the noisy speech and the clean speech, the training data has to be constructed artificially by adding noise to clean speech data. For speech data, we use the McGill TSP speech database1 (French and English) and the NTT Multi-Lingual Speech Database for Telephonometry2 (21 languages). Various sources of noise are used, included computer fans, office, crowd, airplane, car, train, construction. The noise is mixed at different levels to produce a wide range of signal-to-noise ratios, including clean speech and noise-only segments. Since we do not use cepstral mean normalization, we ensure robustness against variations in frequency responses by filtering each of the noise and speech signal independently using a second order filter of the form

$$\boldsymbol{H(z) = \frac{1 + r_1 z^{-1} + r_2 z^{-2}}{1 + r_3 z^{-1} + r_4 z^{-2}}} \qquad (7)$$

where each of $\boldsymbol{r_1}$ ... $\boldsymbol{r_4}$ are random values uniformly distributed in the $\left[-\frac{3}{8}, \frac{3}{8}\right]$ range. Robustness to the signal amplitude is achieved by varying the level of the mixed signal. We have a total of 6 hours of speech and 4 hours of noise data, which we use to produce 140 hours of noisy speech by using various combinations of gains and filters and by resampling the data to frequencies between 40 kHz and 54 kHz.

3.2 Optimization process

The loss function used for training determines how the network weighs excessive attenuation versus insufficient attenuation when it cannot exactly determine the correct gains. Although it is common to use the binary cross-entropy function when optimizing for values in the [0, 1] range, this does not produce good results for the gains

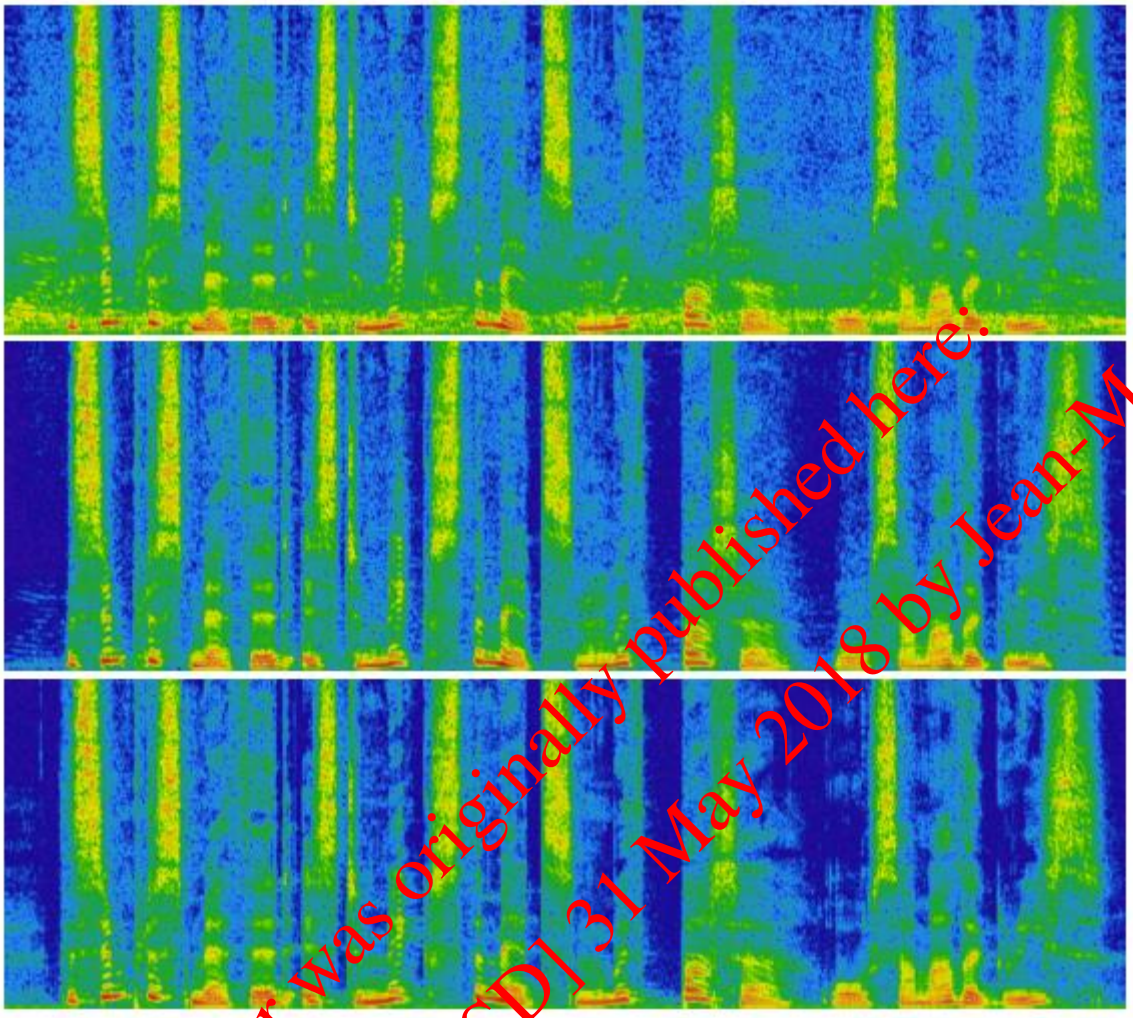

Fig. 4. Example of noise suppression for babble noise at 15 dB SNR. Spectrogram of the noisy (top), denoised (middle), and clean (bottom) speech. For the sake of clarity, only the 0-12 kHz band is shown.

because it does not match their perceptual effect. For a gain estimate $\hat{g}_b$ and the corresponding ground truth g_b, we instead train with the loss function

$$L(g_b, \hat{g}_b) = \left(g_b^\gamma - \hat{g}_b^\gamma\right)^2, \tag{8}$$

where the exponent γ is a perceptual parameter that controls how aggressively to suppress noise. Since $\lim_{\gamma \to 0} \frac{x^\gamma}{\gamma} = \log(x)$, $\lim_{\gamma \to 0} L(g_b, \hat{g}_b)$ minimizes the mean-squared error on the log energy, which would make the suppression too aggressive given the lack of a floor on the gain. In practice, the value $\gamma = {}^1/_2$ provides a good trade-off and is equivalent to minimizing the mean squared error on the energy raised to the power ${}^1/_4$. Sometimes, there may be no noise and no speech in a particular band. This is common either when the input is silent or at high frequency when the signal is low-pass filtered. When that happens, the ground truth gain is explicitly marked as undefined and the loss function for that gain is ignored to avoid hurting the training process. For the VAD output of the network, we use the standard crossentropy loss function. Training is performed using the Keras3 library with the Tensorflow4 backend.

3.3 Gain smoothing

When using the gains $\hat{g}_b$ to suppress noise, the output signal can sometimes sound overly dry, lacking the minimum expected level of reverberation. The problem is easily remedied by limiting the decay of $\hat{g}_b$ across frames. The smoothed gains $\breve{\boldsymbol{g}}_{\boldsymbol{b}}$ are obtained as

$$\breve{\boldsymbol{g}}_{\boldsymbol{b}} = \mathbf{max}\left(\boldsymbol{\lambda}\breve{\boldsymbol{g}}_{\boldsymbol{b}}^{(\boldsymbol{prev})}, \hat{g}_b\right), \quad (\mathbf{9})$$

where $\breve{\boldsymbol{g}}_{\boldsymbol{b}}^{(\boldsymbol{prev})}$ is the filtered gain of the previous frame and the decay factor $\boldsymbol{\lambda} = \mathbf{0.6}$ is equivalent to a reverberation ime of 135 ms.

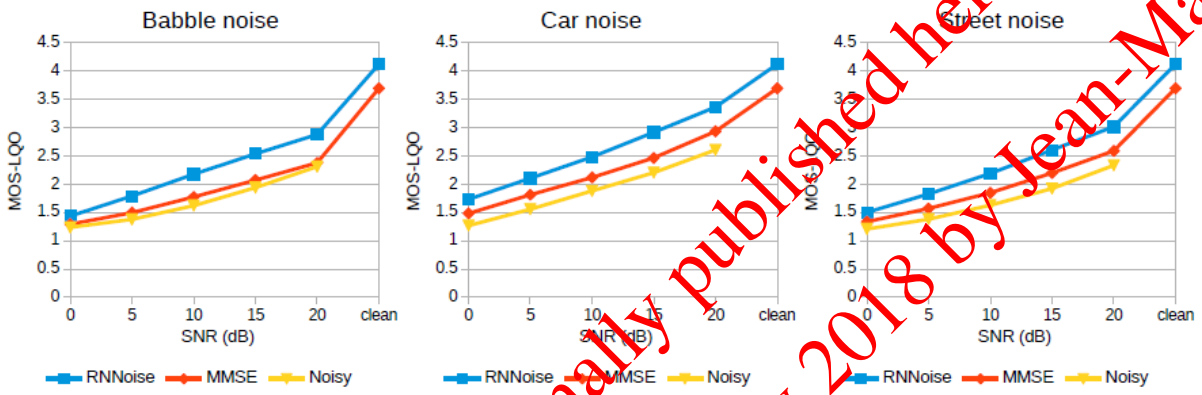

Fig. 5 PESQ MOS-LQO quality evaluation for babble, car, and street noise.

4 Complexity Analysis

We show that the proposed approach has an acceptable complexity and that it provides better quality than more conventional approaches (Sec. 5). We conclude in Sec. 6 with directions for further improvements to this approach. To make it easy to deploy noise suppression algorithms, it is desirable to keep both the size and the complexity low. The size of the executable is dominated by the 87,503 weights needed to represent the 215 units in the neural networks. To keep the size as small as possible, the weights can be quantized to 8 bits with no loss of performance. Since each weight is used exactly once per frame in a multiply add operation, the neural network requires 175,000 floating-point operations (we count a multiply-add as two operations) per frame, so 17.5 Mflops for real-time use. The IFFT and the two FFTs per frame require around 7.5 Mflops and the pitch search (which operates at 12 kHz) requires around 10 Mflops. The total complexity of the algorithm is around 40 Mflops, which is comparable to that of a fullband speech coder. A non-vectorized C implementation of the algorithm requires around 1.3% of a single x86 core (Haswell i7-4800MQ) to perform 48 kHz noise suppression of a single channel. The real-time complexity of the same floating-point code on a 1.2 GHz ARM Cortex- A53 core (Raspberry Pi 3) is 14%.

5 Results

We test the quality of the noise suppression using speech and noise data not used in the training set. We compare it to the MMSE-based noise suppressor in the SpeexDSP library. Although the noise suppression

operates at 48 kHz, the output has to be resampled to 16 kHz due to the limitations of wideband PESQ [15]. The objective results in Fig. 5 show a significant improvement in quality from the use of deep learning, especially for non-stationary noise types. The improvement is confirmed by casual listening of the samples. Fig. 4 shows the effect of the noise suppression on an example.

6 Conclusion

This paper demonstrates a noise suppression approach that combines DSP-based techniques with deep learning. By using deep learning only for the aspects of noise suppression that are hard to tune, the problem is simplified to computing only 22 ideal critical band gains, which can be done efficiently using few units. The coarse resolution of the bands is then addressed by using a simple pitch filter. The low resulting complexity makes the approach suitable for use in video-conferencing systems. We demonstrate that the quality is significantly higher than that of a pure signal processing-based approach. We believe that the technique can be easily be extended to residual echo suppression, for example by adding to the input features the cepstrum of the far-end signal or the filtered far-end signal. Similarly, it should be applicable to microphone array post-filtering by augmenting the input features with leakage estimates like in [16].

7 References

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. on acoustics, speech, and signal processing, vol. 27, no. 2, pp. 113–120, 1979.

[2] H.-G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in Proc. ICASSP, 1995, vol. 1, pp. 153–156.

[3] T. Gerkmann and R.C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 4, pp. 1383–1393, 2012.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 33, no. 2, pp. 443–445, 1985.

[5] A. Maas, Q.V. Le, T.M. O'Neil, O. Vinyals, P. Nguyen, and A.Y. Ng, "Recurrent neural networks for noise reduction in robust ASR," in Proc. INTERSPEECH, 2012.

[6] D. Liu, P. Smaragdis, and M. Kim, "Experiments on deep learning for speech denoising," in Proc. Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[7] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," IEEE Trans. on Audio, Speech and Language Processing, vol. 23, no. 1, pp. 7–19, 2015.

[8] A. Narayanan and D.Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," in Proc. ICASSP, 2013, pp. 7092–7096.

[9] C. Montgomery, "Vorbis I specification," 2004.

[10] J. Princen and A. Bradley, "Analysis/synthesis filter bank design based on time domain aliasing cancellation," IEEE Tran. on Acoustics, Speech, and Signal Processing, vol. 34, no. 5, pp. 1153–1161, 1986.

[11] B.C.J. Moore, An introduction to the psychology of hearing, Brill, 2012.

[12] J.-M. Valin, G. Maxwell, T. B. Terriberry, and K. Vos, "Highquality, low-delay music coding in the Opus codec," in Proc. 135th AES Convention, 2013.

[13] Juin-Hwey Chen and Allen Gersho, "Adaptive postfiltering for quality enhancement of coded speech," IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 59–71, 1995.

[14] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," in Proc. Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8), 2014.
[15] ITU-T, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001.
[16] J.-M. Valin, J. Rouat, and F. Michaud, "Microphone array postfilter for separation of simultaneous non-stationary sources," in Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on. IEEE, 2004, vol. 1, pp. I–221.

Proposal of a Memory Architecture for Pre and Post-Correlation coherent Processing of GNSS Signal with SoC based Acquisition Unit

Jiří Svatoň, František Vejražka
FEE, Czech Technical University in Prague
Prague, Czech Republic

svaton.jiri@fel.cvut.cz
Pavel Kubalík, Jan Schmidt
FIT, Czech Technical University in Prague
Prague, Czech Republic

pavel.kubalik@fit.cvut.cz

Keywords. GNSS acquisition, Parallel search in code algorithm, FFT, Post-correlation coherent processing, Post-correlation coherent processing, Galileo.

Abstract

This contribution describes an architecture of additional system of memories for an existing GNSS (Global Navigation Satellite Systems) signal acquisition unit in frequency domain. The unit is designed for an FPGA-based HW receiver and has three 4K FFT blocks. The receiver is based on the System on Chip (SoC) Xilinx ZYNQ platform. The proposed additional memories are used as accumulators of complex signals samples and are placed in front or after the acquisition unit. They enable to process GNSS signals of different navigation systems more effectively with limited resources.

1 Introduction

GNSS signal acquisition is a process of initial estimation of the following signal parameters: the Doppler frequency shift f_d and the CDMA code phase τ. It is realized by computation of the cross-correlation function of the received signal *r(t)* with a period of CDMA code replica *c(t)* (1). The signal of some modern GNSS signals is quasi-periodic, because it is modulated by the product of primary and so called secondary code. The primary code is a binary CDMA-like code.The secondary code is an additional code with short period (tens of bits). It is applied on whole periods of the primary CDMA ranging code in some types of signals instead of navigation message data. The XOR product of both codes is BPSK modulated to RF signal. The acquisition method, computation of the cross-correlation function in time domain, has quadratic complexity.

There is a faster algorithm called the Parallel-in-Code Search algorithm (PCS) (2) [1], which reduces the complexity to $N\log_2 N$. It computes signal spectra using FFT.

We used the Parallel-in-Code Search for the acquisition unit in an FPGA-based receiver described earlier [2], [3]. The unit has three 4K FFT blocks (Figure 1). It uses noncoherent processing of correlation function results [3] to process signal with a code period longer then the size of its FFT units. This principle was demonstrated on the reception of the Galileo E1B signal. Using a better-suited algorithm and increased size of the FFT units, we reached 3 dB betters results against original [4].

This type of processing of signal divided into partial signal blocks is still limited by the sensitivity loss caused by its noncoherent combining and by partial correlation loss, which is in turn caused by the computation of cyclic-like cross-correlation through frequency domain. Nevertheless, it is still well suited for less efficient systems with latency higher than one signal period or its processed part.

The latency of the discussed unit is less than the processed part of code period of the signal. Therefore, cross-correlation of entire code period could be computed continually in time from its partial correlations and these partial correlations could be then combined coherently to process the signal more effectively. Such a processing, however, deviates from the systolic operation of the original unit and needs additional memories that store both the in-phase (I) and quadrature-phase (Q) signal components.

The architecture of memory added to the existing acquisition unit is presented in Section 2 and 3. The memory is employed as a pre-correlation stage for an effective processing/synchronizing of additional GNSS secondary codes in Section 2. In addition, a method for better processing of long-period code signal avoiding noncoherent combining loss in the post-correlation stage is presented in Section 3. Both principles use an extended coherent time over N periods T_{code} of the primary code. These approaches utilize a linearity of both correlation function in time (1) and frequency domain through FFT (2).

$$\left(\hat{\tau},\, \hat{f}_d\right) = \arg\max_{\tau,\, f_d} \left\{ \left| \sum_{t=0}^{N.T_{code}} r[t]\, c[t-\tau]^{*}\, e^{j2\pi f_d t} \right|^2 \right\} = \arg\max_{\tau,\, f_d} \left\{ \left| R\left(\tau,\, f_d\right) \right|^2 \right\} \qquad (1)$$

$$R\left(\tau,\, f_d\right) = \mathrm{FFT}^{-1}\left\{ \mathrm{FFT}\left\{ r[t]\, e^{j2\pi f_d n} \right\} \mathrm{FFT}\left\{ c[t] \right\}^{*} \right\} \qquad (2)$$

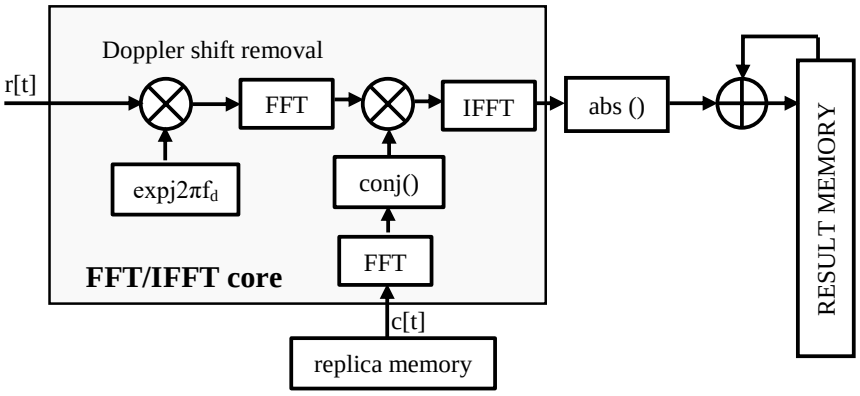

Figure 1. The parallel search in code acquisition diagram with noncoherent combining of results in post-correlation stage

2 Coherent pre-correlation averaging signal processing method

The method was presented in [5] under name mPCA (modified Pre-Correlation Averaging), as a method for secondary code synchronization and secondary code removal. The goal is to obtain better sensitivity for weak GNSS signal scenarios (as is indoor positioning for example) using coherent averaging of signal periods.

This method is assumed to be used before the Parallel-in-Code Search algorithm. Secondary code phase is searched for sequentially. Periods of code are coherently accumulated, respecting the estimated bit of secondary code. A peak of correlation value is detected for an appropriate shift of secondary code.

This type of processing could be realized the same way in post-processing. The main advantage of the pre-processing approach is that coherent accumulation is computed only once per secondary code period. Total saving against a classical post-correlation processing is equal to the number of secondary code bits minus one of cross-correlation computations.

The proposed schema in Figure 2 uses the mPCA principle. Figure 3 shows the Peak to Noise Ratio (PNR) of the correlated signal. The simulation ran over one period of the 20-bit long secondary code of BeiDou B1-I signal and exhibits a dominant correlation peak.

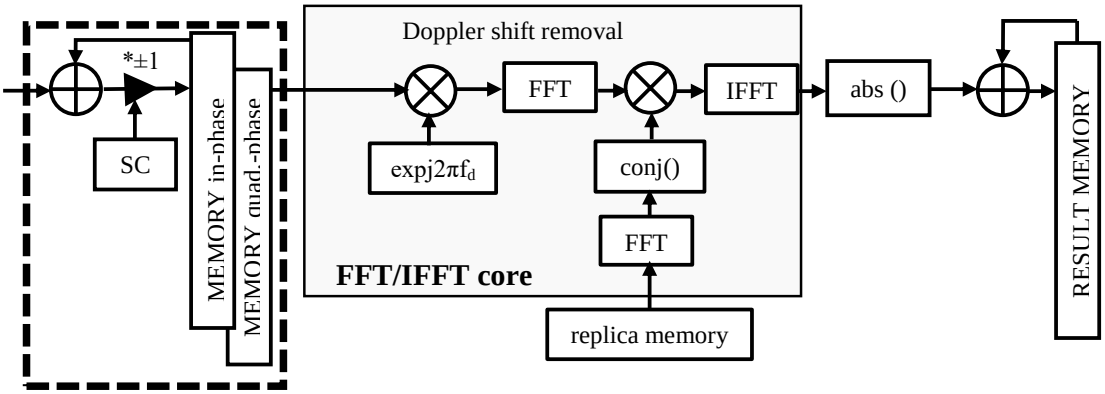

Figure 2. The parallel-in-code search algorithm acquisition unit diagram using PCA processing method

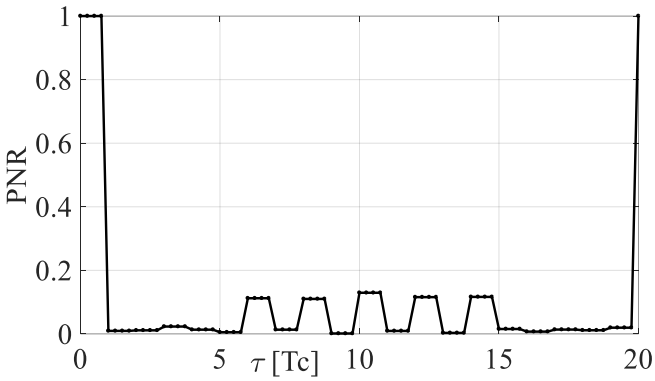

Figure 3. PNR of mPCA acquisition/synchronization of BeiDou B1 secondary code

3 Coherent partial post-correlation signal processing method

This method is intended for the situation where the number of samples N_s of code period T_{code} is much larger than the size N_{FFT} of FFT blocks. The samples, together with the local code replica, are decomposed to M equal-sized consecutive blocks of signal. These blocks of signal samples are partially correlated with the corresponding part of the replica. The results are combined coherently together. The key issue for a coherent combining is to choose M and N_{FFT}, so that each block can be computed with a latency L lower than the signal period divided by M (latency < T/M).

The modification of the original structure in Figure 1 is in Figure 4. The post-correlation stage includes a coherent accumulator with controlled addition/subtraction and a memory for both in-phase and quadrature-phase signal. The controlled accumulator solves the problem of navigation message data bit transition. Every partial correlation could be accumulated with a correct phase for both hypotheses of presence and non-presence of data bit transition.

The gain over a noncoherent processing is equal to 20log *M* in dB. When applied to the Galileo E1B signal, we are able to obtain a 6 dB gain over the original noncoherent method.

4 Summary and future work

The proposed solutions depicted on Figures 2 and 3 allow reusing the core of an existing GNSS acquisition unit. It has been augmented with memories to coherently process signals with long-code period. The unit exhibits better sensitivity in comparison with previous noncoherent combining. It is able to acquire signal having secondary code with a latency up to the length of secondary code. The acquisition unit is a FPGA IP block connected via the AXI bus to the processing system of target SoC Xilinx ZYNQ platform.

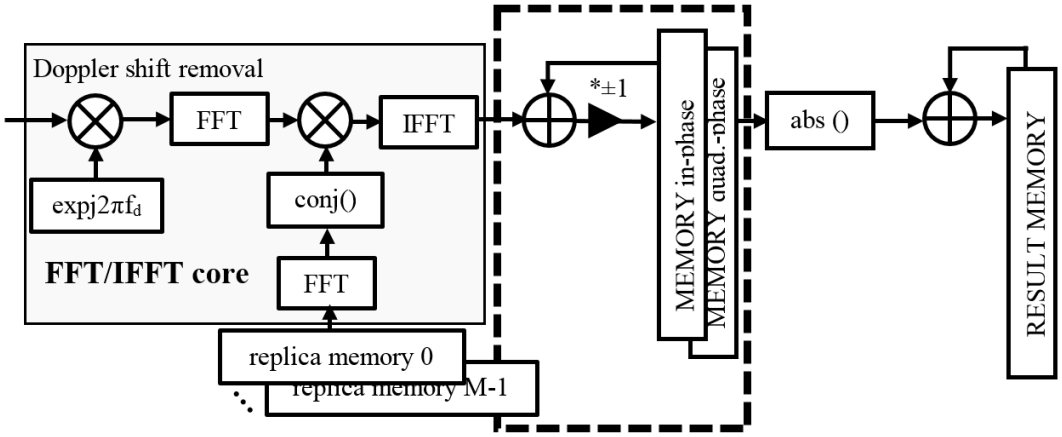

Figure 4. The parallel search in code acquisition diagram with coherent combining of results in post-correlation stage

Paper origin

A part of this paper has been accepted and presented at the Positioning, Location and Navigation Symposium (PLANS) 2018.

Acknowledgment

The submitted paper has arisen during the work on the project, which has been financially supported by the Technology Agency of the Czech Republic under grant No. TE01020186. This work was also supported by the Student Grant Agency of the Czech Technical University in Prague, grant No. SGS18/140/OHK3/2T/13.

References

[1] Van Nee, D.J.R, Coenen, A.J.R.M.: New fast GPS code-acquisition technique using FFT, Electronics Letters 27 (1991): 158-160.

[2] Svatoň, J., Vejražka, F., Kubalík, P., Schmidt, J.: Acquisition of Modern GNSS Signals in SoC ZYNQ with its Limited Computational Resources in Frequency Domain, In: Proceedings of the 5th Prague Embedded Systems Workshop. Praha, 2017. pp. 64-66. ISBN 978-80-01-06178-7.

[3] Svatoň, J, Vejražka, F, Kubalík, P, Schmidt, J.: Methods and Hardware architecture for Multi-constellation GNSS signal acquisition unit in frequency domain, proceedings of European Navigation Conference, May 9-12, 2017, Lausanne, Switzerland, p. 252-261

[4] Fortin, M.-A, Bourdeau, F, Landry, R.Jr.: Implementation Strategies for a Software-Compensated FFT-based Generic Acquisition Architecture with Minimal FPGA Resources, Navigation 62.3 (2015):171-188.

[5] Svatoň, J, VEJRAŽKA F.: Pre- and Post-Correlation Method for Acquisition of New GNSS Signals with Secondary Code. In: IEEE/ION Position Location and Navigation Symposium 2018 (PLANS 2018). Monterey (USA), 23.4.2018 - 26.3.208. in-print.

Application of Neural Networks for Decision Making and Evaluation of Trust in Ad-hoc Networks

Yelena Trofimova, Alexandru Mihnea Moucha, Pavel Tvrdik
Department of Computer Systems, Faculty of Information Technology, Czech Technical University in Prague
Thakurova 9, 160 00 Prague 6, Czech Republic

yelena.trofimova@fit.cvut.cz

Keywords. Ad-hoc network, trust, security, neural network, OMNET++.

Abstract

The lack of infrastructure and central management in ad-hoc networks is an advantage from the viewpoint of scalability and flexibility, but it poses security risks and requires close cooperation among nodes for the network to function as a whole. Nodes have to make a trade-off between saving battery (behaving selfishly) and routing data of other nodes to maintain the network. A part of this trade-off may be the decision of a node to drop some of the data packets in transit that are not either sourced by it or intended for it. Thus, dropped data packets may be a sign of a selfish or even malicious behavior, resulting in performance degradation in the network. Trust-based approach looks promising for improving security and cooperation [1]. The concept of trust in distributed systems arose from the notion of social trust [2]. By the trust problem, we understand the problem of measuring the confidence in the fact that individual node will cooperate - by properly delivering the data in transit, sourced or destined for other nodes. We model trust using the packet delivery ratio (PDR) metric [3].

We have developed a method to apply neural networks (NNs) [4] for solving the problem of trust. It demonstrates that NNs are capable of detection of untrusted nodes and estimation of the trust values. We developed a simulator of ad-hoc networks, containing a generator of datasets. These datasets were used to train and validate NN quality on different data. We have conducted a series of simulation experiments and measured the quality of our method. Our experiments show clearly that NNs can be effectively used for solving the problem of detection of untrusted nodes and trust value estimation.

The results show in average 98% accuracy of the classification and 94% of the regression problem. An important contribution of our research is a verification of the hypothesis that synthetic generation of ad-hoc network traffic in a simulator is sufficient for training of a NN that is then capable to accurately estimate trust. Our NN-based method can be applied in a running ad-hoc network with a given topology. Training of the NNs can be done without collecting data from a running network, since the training data can be constructed artificially. In case of topology changes, new learning of NN can be performed quickly and effectively. No active measurements is needed.

The contributions of this paper are (1) confirmation of applicability of NNs to trust management in ad-hoc networks; (2) construction of a method to detect untrusted nodes and to estimate the value of trust using NNs.

Paper origin

This paper has been accepted and presented at the 13th IEEE International Wireless Communications and Mobile Computing Conference (IEEE IWCMC 2017), held in Valencia, Spain, June 26–30, 2017.

Acknowledgment

This research was supported by grant No. SGS17/214/OHK3/3T/18 from Czech Technical University in Prague.

References

[1] Marchang, N., Datta, R.: Light-weight trust-based routing protocol for mobile ad hoc networks, IET Information Security, vol. 6, no. 2, pp. 77–83, June 2012.

[2] Cook, K.S.: Trust in Society, Russell Sage Foundation Series on Trust, 2003.

[3] Dubey, M., Patheja, P. S., Lokhande, V.: Reputation based trust allocation and fault node identification with data recovery in manet, 2015 International Conference on Computer, Communication and Control (IC4), Sept 2015, pp. 1–6.

[4] Shanmuganathan, S., Samarasinghe, S.: Artificial Neural Network Modelling, 1st ed. Springer Publishing Company, Incorporated, 2016.

Characterizing IP addresses by predicting their malicious behavior

Václav Bartoš
CESNET a.l.e.
Zikova 4, 160 00 Prague, Czech Republic
bartos@cesnet.cz

Keywords. Network security, reputation database, machine learning, attack prediction, probability estimation

Abstract

Security monitoring tools, such as honeypots, IDS, behavioral analysis or anomaly detection systems, generate large amounts of security events or alerts. These alerts are often shared within some communities using various alert sharing systems (such as Warden, AbuseHelper, n6, MISP, *etc*.). Number of alerts processed by such sharing systems may go up to millions per day [1].

In [2] we proposed to build a large reputation database of IP addresses and other identifiers reported as malicious by these alerts to keep track of the potentially dangerous entities on the Internet. Is such a database, each IP address or other entity reported as malicious has a record keeping meta data about all related alerts. The records are further enriched by gathering other information related to the entity, such as its presence on various blacklists, geolocation, information from whois registries, or information about the type of the device (*e.g.* server, mobile phone, IoT, ...) and its connection (*e.g.* dynamic or static IP address).

In this work we show how all the information about malicious entities can be summarized into a small set of numbers, which can be used for quick overview or for ranking by the level of threat each entity poses. More concretely, we define a *future misbehavior probability (FMP) score* as such summarization. It is the probability of receiving another alert about a given malicious IP address within a future time window of specified length (*e.g.* the next 24 hours). The probability is estimated using machine learning methods, taking as input information about previous alerts as well as other data available in the database.

In our method, each sample, *i.e.* an IP address at a specific point in time, is represented by a vector of 56 features, combining information about previous alerts related to the IP address and other IP addresses in the same /24 prefix (since nearby addresses often exhibit similar behavior) with the other information, such as presence of the IP address on several blacklists and various tags derived from the associated hostname, *e.g.* whether the IP address is dynamically or statically assigned, or whether it is a NAT or VPN.

We passed these data to different machine learning models, neural networks (NN) and gradient boosted trees (GBT, also known as xgBoost), with various configurations. In each case the models are fit with the goal of minimizing Brier score – mean squared difference between the predicted probability of a future alert and the actual presence of the alert (labeled as 0 or 1).

For evaluation we took data from an alert sharing system Warden[1] and a reputation database NERD[2], both developed and operated by CESNET. We used two datasets, differing in the type of alerts predicted

[1] https://warden.cesnet.cz/
[2] https://nerd.cesnet.cz/

– network scanning (*scan* dataset and attempts of unauthorized access, *e.g.* by brute force password guessing or exploiting a vulnerability (*access* dataset).

With both datasets, the best performing model is a GBT with 200 decision trees, each of maximum depth of 7. Its Brier score is 0.0638 for *scan* data, 0.0511 for *access* data. By the means of ROC curves, its area under curve (AUC) is 0.9323 and 0.8731 for *scan* and *access* data, respectively. This means the model can indeed predict future attacks and estimate their probability quite well.

The probability (FMP score) can be used to rank IP addresses by the level of threat they pose (by predicting a specific type of malicious traffic only, the threat level may be evaluated in a specific context). This can be used, for example, to create predictive blacklists of any desired size by simply taking top-N addresses with the highest score. The optional size is useful when the blacklist is to be used in a traffic blocking device with a limited maximum number of blocking rules. The FMP score can also be used as a decision criterion in many other algorithms, *e.g.* in alert prioritization (higher priority if the related IP address has high probability of further attacks) or DDoS mitigation (block traffic from all IP addresses with bad reputation, *i.e.* high FMP score). It also servers well as a quick overview of IP address properties for a human analyst.

Note: The work summarized in this abstract is currently under review for the Future Generation Computer Systems journal.

References

[1] Bartoš, V.: Analysis of alerts reported to Warden. Tech. Rep. 1/2016, CESNET

[2] Bartoš, V., Kořenek, J.: Evaluating Reputation of Internet Entities. In: Management and Security in the Age of Hyperconnectivity: Proceedings of the 10th IFIP WG 6.6 International Conference on Autonomous Infrastructure, Management, and Security (AIMS 2016), Springer, 2016.

Grouping evil IP addresses

Lenka Stejskalová, Tomáš Čejka
CTU in Prague
Thákurova 9, 160 00 Prague, Czech Republic

stejsle1@fit.cvut.cz, cejkato2@fit.cvut.cz

Keywords. botnet, distributed attacks, security incident, NERD, Warden, IDEA, Python

Abstract

Botnet is a group of devices that synchronously performs distributed attacks. Botnets currently represent a very dangerous potential threat to all systems. Botnets can attack with great force, especially when it comes to botnets of many hundreds and thousands of bots. Defense against distributed attacks plays an important role in defending the entire system. The Intrusion Detection System (IDS) is part of the defense. This system monitors network traffic and detects suspicious activity that could lead to a system security breach. IDS is a source of reported detected security events that are solved by Incident response. Sharing information from these reports can help you get a global view. The NERD system was developed CESNET [1], the operator of the Czech National Research and Education Network (NREN). System NERD collects information about all malicious entities on the network and manages a reputation database over them.

The motivation to design a system for tracking suspicious network addresses was to extend a set of tools to record suspicious entity information. This system has the task of keeping information about security event records and from this information creating groups of network addresses from events of very similar characteristics. The system creates an evidence of suspicious network addresses.

This presentation describes the design and implementation of the new system GRIP (Group of IPs). The presentation also describes the analysis of security incidents records in IDEA format [2]. Based on this analysis an algorithm was designed to create groups of suspicious network addresses from a security incident.

Acknowledgment

This work was supported by the CTU grant No. SGS17/212/OHK3/3T/18 funded by internal grant of CTU in Prague.

References

[1] CESNET, z.s.p.o.: Network Entity Reputation Database (NERD), https://nerd.cesnet.cz/, last visited 2018-01-18.

[2] CESNET, z.s.p.o.: Intrusion Detection Extensible Alert, https://idea.cesnet.cz/en/index, last visited 2018-03-22.

Fault Tolerance in HLS for the Purposes of Reliable System Design Automation

Jakub Lojda, Zdeněk Kotásek
Brno University of Technology, Faculty of Information Technology, Centre of Excellence IT4Innovations
Bozetechova 1/2, 612 66 Brno, Czech Republic

{ilojda, kotasek}@fit.vutbr.cz

Keywords. Fault Tolerance, High-Level Synthesis, Catapult C, Electronic Design Automation, Robot Controller, C++.

Abstract

Nowadays, many important processes are controlled by electronic systems. Nevertheless, if such system fails, the resulting damage might result in high economical loss or even endanger human health. As an example, autonomous vehicles, which are very popular these days, may serve. Another example might be a device that is possibly not serviceable for a very long period of time, such as a space probe or an artificial satellite. The reparation cost of these devices is well worth the effort to make these devices reliable as much as possible. This effort makes designers focus on the aspects of system reliability. Furthermore, as the chip-level integration grows, the resulting systems are increasingly prone to their failure, further increasing the importance of such aspect of reliability. The complexity of today's modern systems, however, makes this a difficult task to solve.

One possible solution to ensure higher reliability is to make a system so-called Fault Tolerant (FT), which means its ability to perform its function even during presence of faults. The system, however, still remains composed of non-reliable parts. Each method of FT is characterized by the way the non-reliable parts are composed to improve the overall reliability of the resulting system. The solution to the high complexity of today's systems is to move the development to a higher level of abstraction. High-Level Synthesis (HLS) is becoming popular for allowing further move to a higher level of abstraction. Using HLS it is possible to transfer an algorithm (e.g. described in a higher-level programming language, such as C++) to its Register Transfer Level (RTL) representation (e.g. described in a VHDL language).

Our research focuses on an combination of HLS and FT, as we believe the combination of these approaches solves both the problems. We developed an approach to insert reliability to HLS-generated systems. Many approaches to incorporate FT into HLS exist, however, in our approach, the input algorithm is modified *before* its processing by the HLS tool and, thus, the HLS tool itself does not require any modification. The descriptions of FT methods are then made on the same abstraction level as the language used to describe the algorithm, thus, making them easier to develop and maintain. The source algorithm is separated from the FT description making the code easier to maintain as well. The overview of this approach is shown in Figure 1.

In the presentation, this existing method to insert redundancy into HLS-generated systems will be briefly described alongside with its improvement in the form of a majority function selection. We found out that the type of the majority function affects not only the resulting reliability, but also the resources

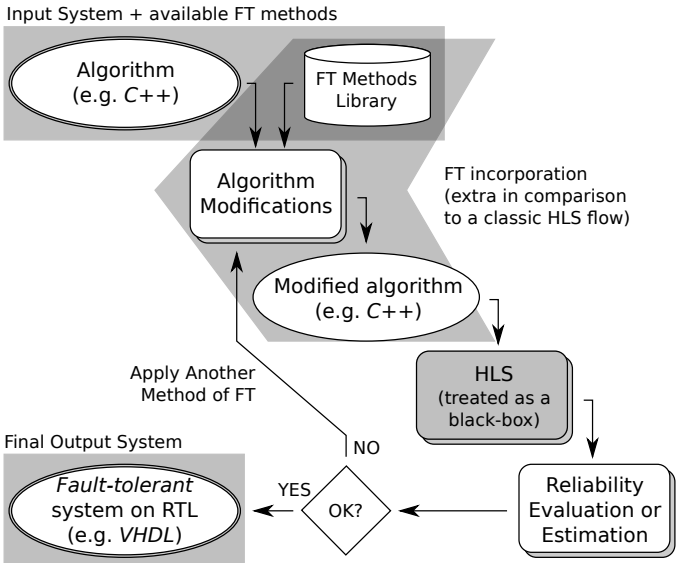

Figure 1: The approach to insert redundancy into systems generated using HLS.

consumption. The presentation also addresses the level of redundancy selection, as we evaluated various numbers of redundant modules with multiple fault occurrences. The case study experiments are carried out with our robot verification platform utilizing the so-called left-hand algorithm and fault injection into a Field Programmable Gate Array (FPGA) implementing the robot controller. This approach is not limited to FPGAs, however, we use an FPGA technology during the evaluation for its wide range of applications and its versatility.

Paper origin

The original paper has been accepted and presented at the 16th IEEE East-West Design & Test Symposium (EWDTS-2017) in Serbia [1].

Acknowledgment

This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II), the project IT4Innovations excellence in science – LQ1602 and the BUT project FIT-S-17-3994.

References

[1] J. Lojda, J. Podivinsky, Z. Kotasek, and M. Krcma, “Data Types and Operations Modifications: A Practical Approach to Fault Tolerance in HLS,” in *IEEE East-West Design Test Symposium (EWDTS)*, Sept 2017, pp. 273–278.

Testing Fault Tolerance Properties: Soft-core Processor-based Experimental Robot Controller

Jakub Podivinsky, Zdenek Kotasek
Brno University of Technology, Faculty of Information Technology, Centre of Excellence IT4Innovations
Bozetechova 2, 612 66 Brno, Czech Republic

{ipodivinsky, kotasek}@fit.vutbr.cz

Abstract

Various electronic systems play an important role in our everyday lives. Some of them serve for fun or to make our lives easier. These systems are useful but not necessary; when they malfunction, the consequences are not critical. On the other hand, there are systems which are more or less critical, and their failure can cause undesirable consequences. For example, a failure in medicine, aviation, the army or automotive systems can cause high economic losses and/or endanger human health. These systems must be protected against the impact of faults, and flawless operation must be ensured. Fault tolerance is one of the techniques that will ensure this. There are many fault-tolerance methodologies targeted towards various systems and technologies, and new methodologies are being investigated. There have been many fault-tolerant methodologies inclined, among others, to *Field Programmable Gate Arrays* (FPGAs) developed and new ones are under investigation, because FPGAs are becoming more popular due to their flexibility and re-configurability. The second reason why so many techniques are inclined to FPGAs is their sensitivity to faults and ability to be reconfigured in the case of fault occurrence. The configuration of FPGAs is stored as a *bitstream* in SRAM memory. The problem is that FPGAs are quite sensitive to faults caused by charged particles. This particle can induce inversion of a bit in bitstream and this may lead to a change in its behaviour. This event is called *Single Event Upset* (SEU).

It is also important to verify these techniques. An evaluation platform for testing fault-tolerance methodologies targeted towards SRAM-based FPGAs (Field Programmable Gate Arrays) was presented and demonstrated in our previous work. Our evaluation platform is based on *Functional Verification*. The main task of functional verification is to check whether a verified circuit meets its specifications. It compares the outputs of a verified circuit running in an RTL simulator with those of a reference model. In the case of the fault injection, the verified circuit must be implemented into the FPGA, so we do not use classical simulation-based functional verification, but modified FPGA-based functional verification. Our platform uses functional verification as a tool for monitoring the impacts of faults injected into an electronic controller implemented into the FPGA. The use of an FPGA development board where an electronic controller is implemented allows us to inject faults directly into the FPGA. A robot for seeking a path through a maze and the processor-based robot controller serve as an experimental system case study. Experimental results with the unhardened and hardened versions of the new processor-based robot controller are presented and discussed.

Two different strategies of fault injection are used in these experiments: *Multiple faults* and *single faults*. Experiments are done for the unhardened version and the TMR version of the processor-based robot controller. The number of verification runs that were performed for each version of the robot controller and each fault injection strategy is 5000 verification runs. Experimental results are compared with the same experiments with the original hard-coded robot controller.

The experimental results for multiple fault injection strategy are summarized in Table 1. It shows the results of both the unhardened and the TMR versions of the processor-based robot controller and it contains a comparison with the original hard-coded robot controller. One can see that the unhardened electronic version failed in 44.02% and the TMR version failed in 8.14% of the cases. This confirms that TMR is a beneficial approach, even though the increase in resource consumption is high. The table also shows the impact of faults on the mechanical robot; a large number of electronic failures leads to the robot stopping in a place which is less critical than a collision with a wall. In comparison with the original hard-coded robot controller, the processor-based robot controller is more susceptible to faults. This fact is evident both for the unhardened and the TMR version. This phenomenon was expected, because the processor represents a more complex design with lots of partial components. These experiments confirmed our expectations.

Table 1: A comparison of the impact of *multiple* faults injected into the unhardened and hardened versions of the processor-based robot controller and the original hard coded robot controller.

Monitored impact	*Processor-based RC*		*Original hard-coded RC*	
	noft	*tmr*	*noft*	*tmr*
Electronic OK [−]	2751	4593	3544	4839
Electronic failed [−]	2201	407	1456	161
Electronic failed [%]	44.02%	8.14%	29.12%	3.22%
Finish not reached [−]	2179	403	1429	161
Collision with wall [−]	55	7	11	0
Robot stop on place [−]	2124	396	1418	161
Reliability improvement [%]	81.5%		88.9%	

As a future work, we plan to apply some sophisticated fault tolerance techniques on the presented experimental electro-mechanical system and repeat the complete evaluation process. One of the possible improvements is the use of reconfiguration for faulty module recovery and synchronization of the recovered module (processor in our case study) with failure-free modules.

Paper origin

The original paper has been accepted for presentation at Euromicro Conference on Digital System Design (DSD 2018) in Prague [1].

Acknowledgement

This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II); project IT4Innovations excellence in science - LQ1602 and the BUT project FIT-S-17-3994.

References

[1] J. Podivinsky, J. Lojda, O. Cekan, R. Panek and Z. Kotasek. Evaluation Platform for Testing Fault Tolerance Properties: Soft-core Processor-based Experimental Robot Controller. Accepted to the 21th Euromicro Conference on Digital System Design. Prague, 2018.

Triple Modular Redundancy Used in Field Programmable Neural Networks

Martin Krcma, Richard Panek, Zdenek Kotasek
Brno University of Technology, Faculty of Information Technology, Centre of Excellence IT4Innovations
Bozetechova 2, 612 66 Brno, Czech Republic

{ikrcma, ipanek kotasek}@fit.vutbr.cz

Abstract

The artificial neural networks are one of the important models of softcomputing and artificial intelligence. They are structures composed of *neurons* interconnected by weighted *synapses*. Basically, the goal of the networks is to learn the relation between two sets of data vectors, to generalize the relation, to determine its features and to use it for the determining the relation of the unknown vectors belonging to the same problem. This capability can be used for classification tasks, for time series and functional prediction, to control tasks, to image recognition, clustering and other tasks.

The implementation of neural networks is challenged with two great neural networks complexities - space complexity and time complexity. The usual solution of both is to use a powerful hardware, such as graphical processor units or processor clusters, which suffer from a high power consumption. For some networks, FPGAs can be one of the possible solutions if a lower power consumption is desired. In this case, the time complexity is solvable by parallelism which is easy to achieve in both FPGAs and neural networks since both are parallel by their nature. The space complexity is bigger problem since an FPGA has limited resources. Thus, there is a need for such designs that exploit the neural networks parallel character for fast computations and save the FPGA resources as well. A Field Programmable Neural Networks (FPNN) concept can be seen as one of the possible solutions.

The concept of FPNNs [4] is meant to simplify the implementation of artificial neural networks in FPGAs by adjusting their properties to be more suitable for implementation into them. The simplification originates from its main feature - a highly customizable structure which makes it possible to establish resource sharing between the original synaptic connections of the neural network. The FPNNs are composed of dedicated interconnected units called neural resources which approximate the original neurons and synaptic interconnections. The units of the first type are called *activators* and represent the original neural network neurons. The other units are called *links* and serve as an approximation of the original synaptic interconnection. Every link disposes of a set of affine operators serving as an approximation of the original synaptic weights.

The goal of this paper is to describe the implementations of FPNNs, both simple and Triple Modular Redundancy (TMR) secured and compare their FPGA resources utilizations.

The VHDL implementation of both types was created according to the original design and schematic [4]. Both, activators and links were designed as separated units communicating with signals. The communication is based on the asynchronous *request - acknowledgement* model. Every neural resource generates requests for all units directly connected to its output (successors) when its computation is done. Once a successor starts to process the request, it sends the acknowledgement back to the original resource. When the original resource receives acknowledgements from all successors, it selects a new input request to process, sends the acknowledgement and begins the computation. The activators also send a *flag* together with the requests. The flag is a constant activator number and it is used in links to select the proper weight to multiply the input data width. The links then propagate the flag to all connected links.

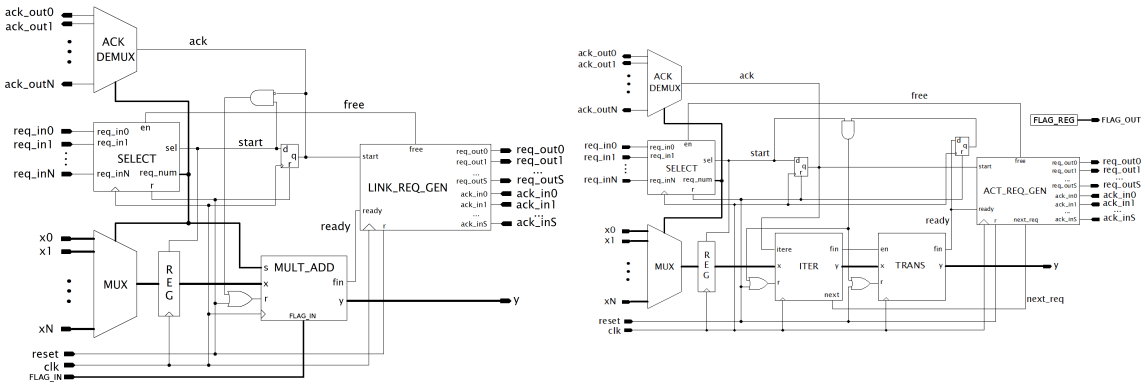

Figure 1: Diagram of a link (left) and an activator (right) implementation - the interconnection of the building blocks

The implementations of both types of neural resources are similar, however they differ in used computational units. The diagram of both types is illustrated in Fig. 1. Both types are composed of a multiplexor, demultiplexor, register, computation units and units for processing requests.

We implemented two versions of TMR secured neural resources and compared their FPGA resources utilizations.

Paper origin

The original paper has been presented at the IEEE East-West Design & Test Symposium 2017 [1].

Acknowledgement

This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II); project IT4Innovations excellence in science - LQ1602, ARTEMIS JU under grant agreement no 641439 (ALMARVI) and BUT project FIT-S-14-2297.

References

[1] M. Krcma and Z. Kotasek and J. Lojda: Triple modular redundancy used in field programmable neural networks. In *IEEE East-West Design Test Symposium 2017*, Sep 2017, ISBN 978-1-5386-3299-4, pp. 1–6.

[2] KRCMA Martin, KASTIL Jan a KOTASEK Zdenek: *Mapping trained neural networks to FPNNs.* In: IEEE 18th International Symposium on Design and Diagnostics of Electronic Circuits and Systems. Belgrade: IEEE Computer Society, 2015, pp. 157–160. ISBN 978-1-4799-6779-7.

[3] Krcma, M.; Kotasek, Z.; Kastil, J.: Fault tolerant Field Programmable Neural Networks. In *Nordic Circuits and Systems Conference (NORCAS): NORCHIP International Symposium on System-on-Chip (SoC), 2015*, Oct 2015, ISBN 978-1-4673-6576-5, pp. 1–4.

[4] Girau, B.: FPNA: Concepts and Properties. In *FPGA Implementations of Neural Networks*, edited by A. R. Omondi; J. C. Rajapakse, Springer US, 2006, ISBN 978-0-387-28487-3, p. 71–123, 10.1007/0-387-28487-7-3.

Stream-wise Aggregation of Flow Data

Michal Slabihoudek, Tomáš Čejka
CTU in Prague
Thákurova 9, 160 00 Prague, Czech Republic

slabimic@fit.cvut.cz, cejkato2@fit.cvut.cz

Keywords. NEMEA, aggregation module, stream-wise processing, UniRec.

Abstract

Network monitoring, especially in large networks, uses so-called flow data analysis. Such analysis is based on aggregation of network packets into IP flows that represent unidirectional communication between pairs of IP addresses. Authors of [1] presented a unique approach to the analysis to handle high data volume of the flow data at near real-time. It is based on a continuous on-the-fly analysis, without permanent storage. Naturally, this approach requires a particular design of the analysis tools. NEMEA [2] is the existing open source detection system that was developed by CESNET, the operator of the Czech National Research and Education Network (NREN), in cooperation with Czech universities. NEMEA uses a UniRec data format that allows for a representation of fixed-sized and variable sized data fields.

There are many NEMEA modules, but a universal aggregation module for the NEMEA system was missing. That is why this work focused on the development of a new NEMEA module that can fulfill the requirements. This presentation describes the design and implementation of the new NEMEA aggregation module. The design was optimized to create a high-performance processing module since it must process a high volume of flow data with a low delay.

The presentation also describes several use cases of the developed module, i.e., connections to existing other NEMEA modules or tools. Finally, the functionality and the performance of the developed module were evaluated, and the presented results confirm that the module is suitable for deployment in monitoring systems of high-speed networks.

Acknowledgment

This work was supported by the CTU grant No. SGS17/212/OHK3/3T/18 funded by internal grant of CTU in Prague.

References

[1] Cejka, T. et al.: NEMEA: A framework for network traffic analysis, CNSM2016, Montreal

[2] CESNET, z.s.p.o.: NEMEA, https://github.com/CESNET/NEMEA, [05. 2018].

Stream-wise adaptive blacklist filter based on flow data

Filip Šuster, Tomáš Čejka

FIT, CTU in Prague
Thákurova 9, 160 00 Praha 6

sustefil@fit.cvut.cz, cejkato2@fit.cvut.cz

Keywords: NEMEA, blacklist filter, IP flow, detection

Abstract

The Internet is full of activists with malicious intentions. Ones tend to steal users' data, others blackmail users for ransom. Luckily, there are projects fighting malicious users and malware in general, for example, by providing public blacklists. Network security initiatives like abuse.ch provide a wide range of blacklists covering different types of malicious activities like botnets, phishing etc. In the network analysis system called NEMEA[1], which is an open source IDS developed by CESNET[2], we are currently focusing on such detection using these publicly available blacklists.

The NEMEA system operates with IP flow data. A flow is an aggregation of network packets and represents an unidirectional IP connection between two endpoints. These flows can be extended with application layer information (L7) such as HTTP or DNS. Simple blacklist detection seems straightforward, i.e. inspecting every IP flow for blacklisted IP addresses, domain names or URLs and reporting this incident to Warden (system for sharing detected events). Our detector tries to go beyond that using so called adaptive filter. This filter dynamically enriches the blacklists with additional records by observing patterns in the detected communication. The presentation focuses on examples of these patterns and scenarios where such adaptivity could raise the detection effectiveness.

Below is a picture of the high-level detection architecture, where Adaptive filter controller contains the logic of analyzing patterns and adapting the filter rules. Evaluator then searches for interesting scenarios in the detected traffic.

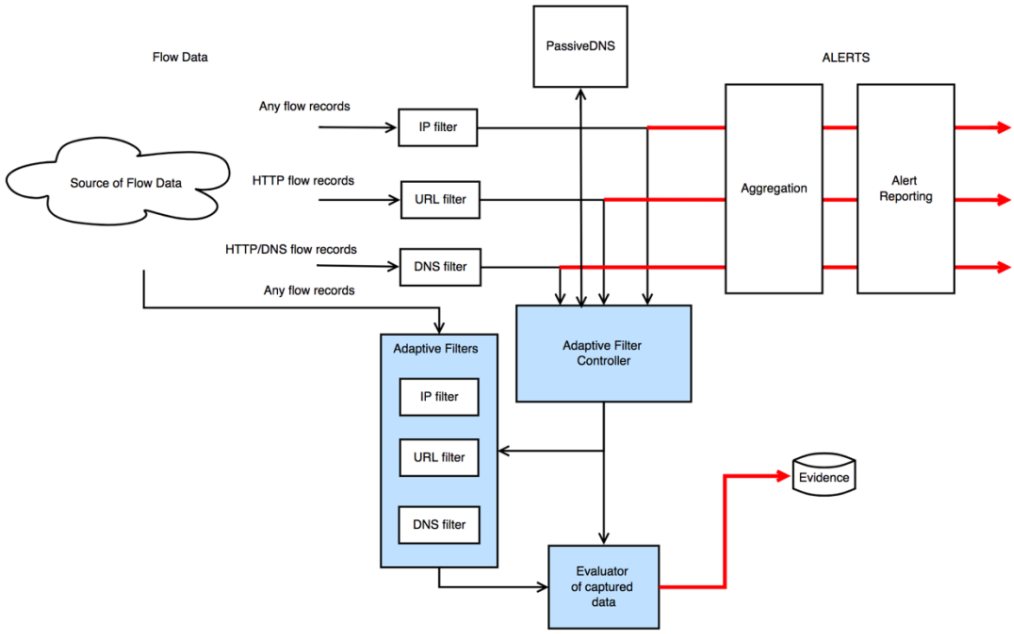

Figure 1: Blacklist detection architecture

Acknowledgment

This was supported by the CTU grant No. SGS17/212/OHK3/3T/18 funded by internal grant of CTU of Prague.

References

[1] Cejka, T. et al.: NEMEA: A framework for network traffic analysis, CNSM2016, Montreal
[2] CESNET, z.s.p.o.: NEMEA, https://github.com/CESNET/NEMEA

Penetration Testing & Web Application Intrusion Detection

Tomáš Ďuračka
ESET Research Czech Republic s.r.o.
Jankovcova 1037/49, 170 00, Prague 7 - Holešovice

tomas.duracka@eset.cz

Keywords. Penetration Testing, Web Application Intrusion Detection, NEMEA, IDS, WAF, OWASP ModSecurity CRS

Abstract

For the last several years, we have been observing a rapid growth in the amount of web application attacks, which resulted in several confirmed sensitive information leakages and led to many web servers becoming hosts for various malicious activities. There are several security measures the companies use to protect themselves from web application attacks. The presentation introduces you to the basics of two of them. Web application penetration testing and web application intrusion detection based on network flows as a part of CESNET NEMEA system [1].

References

[1] Cejka, T.; Bartos, V.; Svepes, M.; aj.: NEMEA: A Framework for Network Traffic Analysis. In 12th International Conference on Network and Service Management (CNSM 2016), 2016.

Informed DDoS Mitigation at 100 Gb/s

Tomáš Jánský, Tomáš Čejka, Martin Žádník, Václav Bartoš
CTU in Prague, CESNET a.l.e.
{Thákurova 9, Zikova 4}, 160 00 Prague, Czech Republic
{janskto1,tomas.cejka}@fit.cvut.cz,{zadnik,bartos}@cesnet.cz

Keywords. amplification DDoS, DDoS mitigation, reputation score, traffic filtering

Abstract

Network attacks, especially DoS and DDoS attacks, are a significant threat to all providers of services or infrastructure. The most potent attacks can paralyze even large-scale infrastructures of worldwide companies (as it is mentioned, e.g., in [1]). The objective of DDoS attacks is usually to flood the target network device or even the network itself with a large number of packets. Such attack results in nondeterministic discarding of network packets.

There are many different types of DDoS attacks hence every mitigation technique addresses only a portion of them. Network operators can use various ways of defense (such as blackholing, rate-limiting) that deterministically discard packets of the traffic according to defined rules. The problem of packet discarding is related to the availability of the victim. When all packets targeted against the victim are discarded, the attack becomes harmless. Naturally, legitimate packets are discarded as well. Therefore, it is not always a feasible approach in practice.

The main challenge is to distinguish malicious and legitimate packets. DDoS mitigation strategy based on the recognition of malicious packets is a complex task due to the similarity between legitimate and malicious packets. This presentation proposes a design of a mitigation heuristic which utilizes the knowledge of the so-called *reputation score* [2] of network entities and describes a way to integrate the proposed heuristic into a scrubbing center developed by *CESNET a.l.e*.

The result, which will be described in this presentation, is based on the DDoS Mitigation Device (DMD) [3] that works at link speed 100 Gb/s. The DMD analysis the traffic on-the-fly, it computes statistics and using our proposed heuristic algorithm based on reputation scores it determines what packets to discard.

Acknowledgment

This was supported by the CTU grant No. SGS17/212/OHK3/3T/18 funded by internal grant of CTU of Prague.

References

[1] Krebs, B. DDoS on Dyn Impacts Twitter, Spotify, Reddit
https://krebsonsecurity.com/2016/10/ddos-on-dyn-impacts-twitter-spotify-reddit/

[2] Bartoš, V., Kořenek, J.: Evaluating Reputation of Internet Entities. In: Management and Security in the Age of Hyperconnectivity: Proceedings of the 10th IFIP WG 6.6 International Conference on Autonomous Infrastructure, Management, and Security (AIMS 2016), Springer, 2016.

[3] CESNET, a.l.e. DDoS Protector, https://www.liberouter.org/technologies/ddos-protector/

P4-to-VHDL: How We Built the Fastest P4 FPGA Device in the World

Pavel Benáček
CESNET, a.l.e.
Zikova 4, 160 00 Prague 6

benacek@cesnet.cz

Keywords. P4, FPGA, High-Level Synthesis, 100Gbps

Abstract

The P4 language [1] is a general platform agnostic language used for the description of packet processing functionality. So far, it is being supported by a huge number of technological leading companies like Google, Intel and so on. The language itself is the next step in the evolution of Software-Defined Networking (SDN) [2]. The SDN concept provides a way for fast deployment of new services into the existing infrastructure due to the high reconfigurability of the SDN ready devices. The most popular embodiment is the OpenFlow protocol [3] which allows us to program OpenFlow-ready switches with a user-defined processing chain. However, the OpenFlow devices are still somehow fixed because the set of supported actions and protocols is fixed because the OpenFlow standard strictly defines the set of supported protocol and actions. One possibility how to overcome this limitation is the usage of an FPGA based device which binds together the high-performance and programmability. However, FPGA circuits are programmed using HDL languages which are not easy to learn for novices and the development itself is time-consuming.

To solve the proposed issues, researchers came with the P4 language which is target independent and it is used for the description of packet processing functionality (including the specification of supported protocols and actions). The front-end of the compiler was released as open source project for Python and C++ language. Using the compiler's front-end, we built a high-level synthesis tool from P4 to VHDL which is capable to generate a VHDL description of a high-speed network device capable to hit the throughput ranging from 77.6 to 100Gbps and possibly beyond in single FPGA. This achievement is very important from the practical point of view because we can change the functionality of FPGA based network device with a program in the P4 language which is then translated to VHDL and synthesized using the standard toolchain. The advantage of this approach is also the fact that developer doesn't need to know anything about HDL programming because the translation from a P4 program to bitstream can be automatized.

The SDN concept is tightly connected with Network Function Virtualization (NFV). The NFV concept is based on the idea that data processing can be done with the series of chained actions. Each action (like computation of statistics, traffic labelling and so on) is performed by a NFV node. The NFV node is typically a virtual machine which is deployed in NFV infrastructure. Therefore, this approach allows us to dynamically react to current requirements and it also allows easy and fast deployment of new functions. Another outcome from the approach is the requirement on the NFV infrastructure which has to be highly configurable. Therefore, the usage of SDN approach is highly

desired and the P4 language can push this forward because it adds another degree of flexibility – you can define your own packet processing device which is tailored to needs of a problem.

The presentation will provide a brief overview of mentioned technologies and their connections in the SDN ecosystem. Another part of the presentation will be reserved for the architectural overview of our modular high-speed network device which is generated from provided P4 program using the P4-to-VHDL tool. Currently, it is the fastest available HLS solution for the P4 language.

Acknowledgment

This research has been partially supported by the Technology Agency of the Czech Republic project TH02010214.

References

[1] The P4 Language Consortium. 2017. The P4 Language Specification. (24 May 2017). https://p4lang.github.io/p4-spec/p4-14/v1.0.4/tex/p4.pdf

[2] W. Xia, Y. Wen, C. H. Foh, D. Niyato and H. Xie, "A Survey on Software-Defined Networking," in IEEE Communications Surveys & Tutorials, vol. 17, no. 1, pp. 27-51, Firstquarter 2015. doi: 10.1109/COMST.2014.2330903

[3] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. 2008. OpenFlow: enabling innovation in campus networks. SIGCOMM Comput. Commun. Rev. 38, 2 (March 2008), 69-74. DOI=http://dx.doi.org/10.1145/1355734.1355746

Anomaly Detection in the SIoT Gateway

Dominik Soukup
CTU in Prague
Thákurova 9, 160 00 Prague, Czech Republic

soukudom@fit.cvut.cz

Keywords. IoT, anomaly detection, BeeeOn, NEMEA.

Abstract

The concept of the Internet exists dozens of years and for lots of people it is very important part of daily life that generates a huge business value. Currently, the number of connected devices grows very fast and this increase should still continue in the future. The main reason for this growth is an expansion of a network connection to almost all electronic devices and sensors. As a name for this trend is used the Internet of Things (IoT).

This work is focused on security concerns and issues of the IoT. The first aim is to analyse the actual situation of IoT and to identify vulnerabilities of the wireless sensor network protocols. The second aim is to develop a tool that is able to detect security incidents in communication traffic. The analytical part describes the fog computing concept and new communication architecture. Simultaneously, there are thoroughly explored current IoT protocols including their vulnerabilities. This is followed by the tool design that is ready for the future extension, which is necessary for this rapidly growing area like IoT. During designing, low hardware requirements were emphasised so that it would be possible to deploy the created solution event on IoT gateways with restricted resources.

The first result of this work is research of the current IoT state, which is contained in the text of this work. The second result is a modular system that is configurable and customizable for target topology. The created tool is implemented in C++ language and extends the already existing IoT gateway BeeeOn by anomaly detection of the wireless sensor network protocols. The result is a new version of the BeeeOn gateway with the mechanism for attacks detection.

Acknowledgment

This work was supported by the CTU grant No. SGS17/212/OHK3/3T/18 funded by internal grant of CTU in Prague.

Monitoring network and threats on the Internet with Turris router

Michal Hrušecký
CZ.NIC
Milešovská 1136/5, 130 00 Praha 3
Michal.Hrusecky@nic.cz

Keywords. network, security, opensource, ids, ips, turris

Abstract

At CZ.NIC we are developing open source routers. Those router actually have quite some resources at their disposal like one gig of RAM, dual core ARM CPU and quite some storage. This gives us space to do some interesting stuff with the traffic that is flowing through our router and take some extra measures regarding security.

1.1 Suricata

We try to provide users with an option to get a deeper insight into what is happening in their home network. Big part of this solution is open source software called Suricata. It is intended as IDS/IPS. We run it on our router in a setup where it can be easily used as IPS, but so far we use just IDS functions. We are just scratching the surface of what is possible, but even now we are able to provide our users with the list of devices and what servers they were trying to access including the correct domain names even in SNI case. In the long term, we plan to expand on this functionality and include some blocking as well, but even now, the information we provide to our users are valuable.

1.2 Firewall, Honeypots and Minipots

Other aspect of our secure router is to monitor general threats on the internet to the end users and report them to us. To monitor those we use various methods but we require explicit consent as we can get some sensitive data this way. The simplest method is collecting firewall logs – who tried to access some port on the router from WAN and was rejected. More complex method is something we call minipot – we serve a fake service (telnet, http, …) and we ask just for username and password. Compared to a honeypot, we are not trying to mimic target system, just the login to it. Last but not least, we use honeypots for ssh. But not an ordinary honeypot, but honeypot as a service where our routers are just man in the middle and are forwarding wanna be attacker to our big honeypot. This way, there is no danger to people running it, but we still get full log with all information.

1.3 Outcome

All the information from firewall logs, minipots and honeypots ends up on our servers. Our goal is to identify malicious attempts and alert the routers to block the attackers. We used to collect all those data in a traditional database and run some statistics on top of it. It was fine and quite fast when we started the project with thousand routers and it allowed us to experiment with various approaches to the statistics. It was even able to scale for some time. But with Turris Omnia, we started getting more and more data and traditional database with running statistics periodically on top of it started to be the bottleneck. Therefor nowadays we are rewriting the system to make it more robust and scalable. To do so, we are using ZeroMQ and Microservices. We should be able to handle more data faster and when it is needed we should be able to scale up by duplicating the most busy services.

KETCube – the Prototyping and Educational Platform for IoT

Jan Bělohoubek
Faculty of Electrical Engineering, University of West Bohemia in Pilsen
Pilsen, Czech Republic

belohoub@ket.zcu.cz

Keywords. IoT, Education, Prototyping, KETCube.

Abstract

The IoT (Internet of Things) devices continue to penetrate into new areas of our daily lives as well as industry. The evolution of IoT devices comes with the necessity of operation in heterogeneous environments. This evolution brings new challenges in the areas of R&D and education. We identify important features beneficial for R&D engineers as well as for educationalists and students and we propose a novel open platform for rapid development of IoT nodes. This platform is easy to employ in the educational process at the same time.

2 Motivation

The continuous and fast movement in the application area stimulates the progress in many fields – IoT LPWAN (Low Power Wide Area Network) standardization efforts [1], infrastructure efforts like LoRaWAN, Sigfox, NB-IoT and many more [1, 2].

As the IoT field in general is very heterogeneous [1], developers of physical devices face many challenges coming from this heterogeneity. An example of a significant challenge for IoT nodes is the (in)ability to gain the profit coming from overlapping networks based on different communication standards [2].

Another challenge coming from the developing area of IoT is connected with the technical education [3]. Educationalists all-around the world face the problem how to introduce students to IoT world without missing any important technology while providing detailed technical insight at the same time.

3 KETCube Platform

Based on the IoT design and educational experience of our team, we decided to release our newly developed prototyping and educational platform supporting our R&D process as well as our educational activities under the non-restrictive University of Illinois/NCSA Open Source License [4]. We call our platform KETCube [5]. The name of the platform – KETCube – consists of the abbreviated name of the institution of its origin (Department of Technologies and Measurement, University of West Bohemia in Pilsen) and the shape of the basic sensor node – a cube – see Figure 1.

The current release of the KETCube platform includes the main board, battery board, datasheet, three application notes and firmware (v0.1). All the mentioned project parts allow the out-of-the-box use of KETCube as an Relative Humidity and Temperature sensor node in LoRaWAN or proprietary

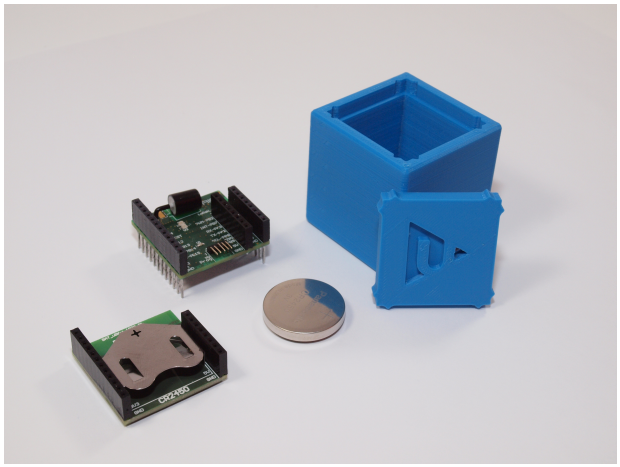

Figure 1: KETCube platform hardware: the main board with a bulk antenna, box, battery board and CR-2450 battery

network. Included documents serve for quick start with prototyping, while providing deep insight to released KETCube parts. Documentation is written in an industry-standard style and such a way serves as a handy guide for in-education deployment at the same time.

The platform has already been used in certain projects and educational activities including the *Object Presence Sensing Demonstrator* and *Enviromental LoRaWAN Sensor Prototype*. Materials related to KETCube are available online on GitHub[1].

4 Conclusions

We identified challenges coming from a heterogeneous and fast growing area of IoT related to the sensor node development. Based on our experience in R&D and educational process, we proposed a novel KETCube platform intended to support of both R&D and educational activities.

References

[1] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low Power Wide Area Networks: An Overview," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 855–873, Secondquarter 2017.

[2] J.-P. Bardyn, T. Melly, O. Seller, and N. Sornin, "IoT: The era of LPWAN is starting now," in *European Solid-State Circuits Conference, ESSCIRC Conference 2016: 42nd*. IEEE, 2016, pp. 25–30.

[3] J. He, D. C.-T. Lo, Y. Xie, and J. Lartigue, "Integrating internet of things (iot) into stem undergraduate education: Case study of a modern technology infused courseware for embedded system course," in *Frontiers in Education Conference (FIE), 2016 IEEE*. IEEE, 2016, pp. 1–9.

[4] "The University of Illinois/NCSA Open Source License (NCSA)," https://opensource.org/licenses/NCSA, accessed: 2018-04-11.

[5] *KETCube datasheet*, University of West Bohemia in Pilsen, 2018, rev. 05/2018, https://github.com/SmartCAMPUSZCU/KETCube-docs/blob/master/KETCubeDatasheet.pdf.

[1] `https://github.com/SmartCAMPUSZCU/KETCube-docs`

Introduction to logic synthesis of polymorphic electronics

Adam Crha
PhD student (Faculty of Information Technology, Brno University of Technology)
Bozetechova 2, Brno, 612 66

icrha@fit.vutbr.cz

Keywords. Polymorphic electronics, polymorphic gates, unconventional, logic synthesis, AIG, PAIG, polymorphic, multiplexing, PolyBDD

Abstract

The most of computer systems are based on inorganic semi-conducting materials also known as silicon. These elements, fabricated from silicon, are called transistors and they are used for building more complex elements - logic gates, which can perform basic boolean functions. This is conventional electronics and is designed by known conventional methodologies. Nowadays, the more interesting technologies are available, that can bring some advantages into the systems where they are applied. It is talked especially about organic semiconductors, semiconductors based on graphene, silicon nano-wires, that can exhibit different behavior depending on the environment state [1]. Polymorphic electronics takes the described behavior as main advantage and the aim of polymorphic electronics is save and share resources in the case of conventional usage or perform any safety behavior in the case of failure. The main open issue consists in polymorphic circuit design. It is too hard design described circuits and conventional synthesis methods are not applicable [2].

Design of polymorphic circuits can be described as finding graph G, which represents interconnection of polymorphic circuit and this circuit performs one of all desired function dependently on the environment state . When a function is switched, interconnection of circuit remains the same, only function of nodes are changing their function [3]. A few synthesis methods were proposed, but all of them have some limitations [4].

This work focuses on the basics of polymorphic electronics and currently known issues in this research area. At the beginning, main purposes and main ideas of the polymorphic electronics will be explained and then the presentation will describe a logic synthesis problem of polymorphic circuits. Already existing synthesis methods will be enumerated and all advantages/disadvantages considered. At the end we look under-hood of current research of new logic synthesis method of polymorphic circuits.

References

[1] H. Raza, “Graphene Nanoelectronics: Metrology, Synthesis, Properties and Applications. ,” 2012.

[2] A. Stoica, R. Zebulum, and D. Keymeulen, “Polymorphic electronics. Proc. of Evolvable Systems: From Biology to Hardware Conference,” vol. 2210 of LNCS, pp. 291–302, 2001.

[3] R. Růžička, “Polymorfní elektronika.” p. 118, 2011.

[4] Z. Gajda, “Evolutionary Approach to Synthesis and Optimization of Ordinary and Polymorphic Circuits ,” 2011.

Hybrid Enhanced Petri Net Model

Almotasem Essa, Zbyněk Jakš
Faculty of Information Technology, Czech Technical University in Prague
Thakurova 9, 160 00 Prague, Czech Republic

essaalmo@fit.cvut.cz

Abstract. Petri nets are powerful formal models. They are based on strict mathematical theories. Petri nets are appropriate for modeling and analyzing systems with parallelization, synchronization and confliction, they provide convenience for qualitative and quantitative analysis in the design phase. A system modeled by a Petri net is easily extended. Petri Nets also provide visual and hierarchical modeling methodologies. Petri nets take care of implementing a wide variety of properties during system design, like safety and security and reliability and they ensure the reachability of the system using multiple predefined constraints, so the system can perform a required task or mission for a specified time in a specified environment. This paper gives ideas how to use such formal model in the whole design process until final realization by hardware and/or software.

Keywords. Petri Nets, PNML, FPGA.

1 Introduction

This paper describes first ideas and proposals for exploitation of formal model – Petri Net for not only modeling of systems, but using them directly for automatized design of digital systems with special constrains which characteristic for embedded systems and IoT.

Several different types of formal methods and models (Petri Nets, Markov chains, UML diagrams) will be used to construct hybrid enhanced model for simplification and automation of digital design process. The selected and/or developed modeling device should meet the requirements that take into account current trends in the field of digital design. It means that they may cover the possibility to model both hardware and software parts of a system. The connection of verification methods and dependability modeling at all design periods to obtain an optimized structure according different parameters must be taken into account. The aim should be the combination of different models and detailed study of their relations and possible automatic conversions to the final implementations. Partial results and proposed methods will be evaluated by real-life applications and benchmarks.

The result should be a hierarchical tool for modeling all dependencies that will be able to express, to verify (functionally and/or formally), to validate, or to certify the proposed system. The system will be specially intended for use in the area of embedded systems used in control systems for critical applications. The aim is to use methods and tools to predict system properties with guaranteed parameters: size, working frequency, power consumption, dependability (reliability, availability, maintainability, safety, fault-tolerance, attack-resistance) and to use partial research results of Digital Design & Dependability Research Group (DDD) in Faculty of Information Technology CTU in Prague.

As the first suitable model Petri Nets have been studied here. We prefer to use some tool available with standardized output format (PNML). We propose the automated translator – compiler from this output description of analyzed and simulated Petri Net model to hardware-software realization.

Therefore the parser PNML2VHDL were designed and tested on two particular and simple examples (dinning philosophers and Produce/Consumer system.

In this paper we will discuss a possible simple realization of mathematically described model (Petri net). The model has to be analyzed and simulated in order to proceed with future research. The current process will explain the process of transforming the chosen model described as a Petri net into synthesizable VHDL code by designing all the necessary blocks of circuits in FPGA. The result of this design will be the presentation of automatic SW tool PNML2VHDL which can handle most of the conversion.

The structure of this paper is as follows: after Introduction the basic descriptions of Petri Net models and PNML properties are described, then the method how to translate PNML output to the hardware implementation of PN and finally the results of one example }dinning philosophers will be presented in chapter Experiments. The last chapter Conclusions gives the future plans of this research.

2 Theoretical background

Petri nets, introduced by C. A Petri in 1962 [54], provide an elegant and useful mathematical formalism for modeling concurrent systems and their behaviors. In many applications, however, modeling by itself is of limited practical use if one cannot analyze the modeled system. As a means of gaining a better understanding of the Petri net model, the decidability and computational complexity of typical automata theoretic problems concerning Petri nets have been extensively investigated in the literature in the past four decades.

Petri nets have been specifically designed to model systems with interacting components and are able to capture many characteristics of an event driven system, namely concurrency, asynchronous operations, deadlocks, conflicts, etc. Furthermore, the PN formalism may be used to describe several classes of logical models (e.g., P/T nets, Colored PNs, nets with inhibitor arcs), performance models (e.g., Timed PNs, Time PNs, Stochastic PNs), continuous and hybrid models (continuous PNs, hybrid PNs)

The main features of PNs can be summarized in the following items.

- PNs are both a graphical and mathematical formalism. Being a graphical formalism, they are easy to interpret and provide a useful visual tool both in the design and analysis phase.
- They provide a compact representation of systems with a very large state space. Indeed they do not require to explicitly represent all states of a dynamical system but only an initial one: the rest of the state space can be determined from the rules that govern the system evolution. Thus a finite structure may be used to describe systems with an infinite number of states

2.1 Definition of Petri Net

A Petri net is formally defined as a 5-tuple N = (P, T, I, O, M0), where

(1) P = {p1, p2, …, pm} is a finite set of places;
(2) T = {t1, t2, …, tn} is a finite set of transitions, $P \cup T \neq \emptyset$, and $P \cap T = \emptyset$;
(3) I: $P \times T \rightarrow N$ is an input function that defines directed arcs from places to transitions, where N is a set of nonnegative integers;
(4) O: $T \times P \rightarrow N$ is an output function that defines directed arcs from transitions to places; and (5) M0: $P \rightarrow N$ is the initial marking.

A marking in a Petri net is an assignment of tokens to the places of a Petri net. Tokens reside in the places of a Petri net. The number and position of tokens may change during the execution of a Petri net. The tokens are used to define the execution of a Petri net.

2.2 PNML

Petri Net Markup Language (PNML) is a proposal of an XML-based interchange format for Petri nets (it is ISO/IEC 15909 standard). Originally, the PNML was intended to serve as a file format for the Java version of the Petri Net Kernel. But, it turned out that currently several other groups are developing an XML-based interchange format too. So, the PNML is only one contribution to the ongoing discussion and to the standardization efforts of an XML-based format.

The specific feature of the PNML is its openness: It distinguishes between general features of all types of Petri nets and specific features of a specific Petri net type. The specific features are defined in a separate Petri Nets Type Definition(PNTD) for each Petri net type.

Furthermore, several specific features are used in more than only one Petri net type. Therefore, there is a conventions documents containing specific Petri net features. Thus, a concrete PNTD adds its its type specific features to PNML by referring to the Conventions Document. The standardization efforts have mainly an effect on this Conventions Document. PNML supports three types of Petri nets, Place/Transition-Nets, High-level Petri Nets a Symetric Nets, see [8]

The following code is an example of how PNML representation look like :

```
<place id="P1">
<graphics>
<position x="525.0" y="15.0"/>
</graphics>
<name>
<value>P1</value>
<graphics>
<offset x="0.0" y="0.0"/>
</graphics>
</name>
<initialMarking>
<value>Default,1</value>
</initialMarking>
<capacity>
<value>1</value>
</capacity>
</place>
```

2.3 Dining philosophers problem

The Dining Philosopher Problem states that K philosophers seated around a circular table with one chopstick between each pair of philosophers. A philosopher may eat if he can pickup the two chopsticks adjacent to him. One chopstick may be picked up by any one of its adjacent followers but not both.

Using standard Petri Net, we can construct a solution by adding places and transitions in a proper way so the transitions will be fired after the token are filled correctly and according to the solution. This is how the design should look like Figures 1 and 2:

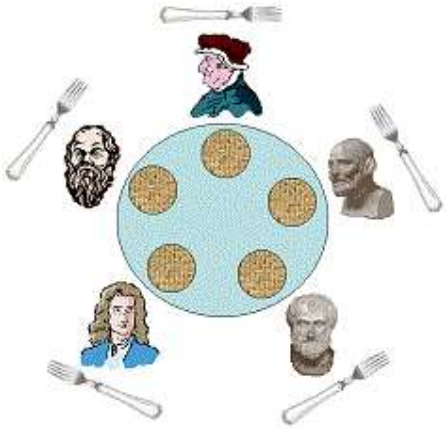

Figure 1 Dining philosophers' representation

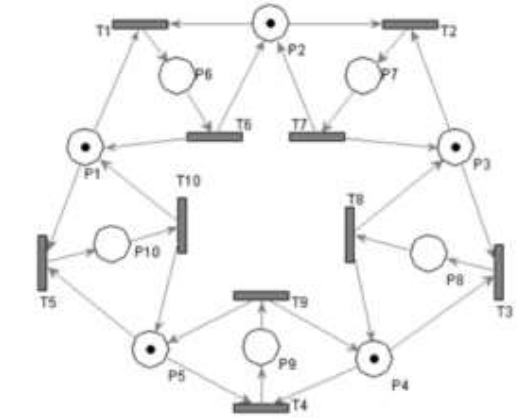

Figure 2 Petri Net solution for dining philosophers' problem

3 FPGA Implementation of a Petri net

The standard Petri Net model was used for proper functionality, and we need some component that controls the flow in PN and guarantees the proper behavior, and this component is called controller, and we also need a component which, according to the definition in case of multiple active transitions, chooses one randomly, which is chooser.

There were several concepts of PN realization as HW as in [1], this thesis explains the realization of PN which supports multiple tokens in place and all other explanations will be based on this idea.

The place (Figure 4) is a counter with a logic for inspecting capacity limit. Since HW is not infinite, there is a limit (255 tokens) after which the place starts showing the infinity flag and stops responding to inc and sub inputs

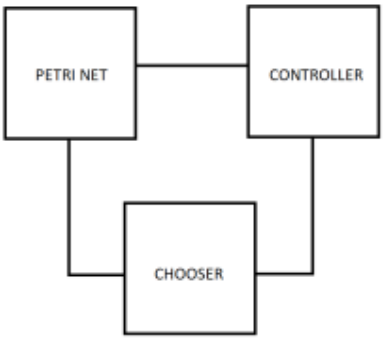

Figure 3 FPGA design of Petri Net

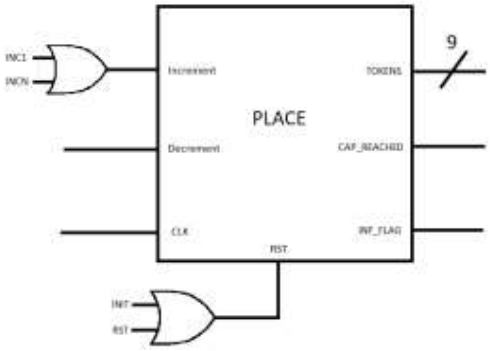

Figure 4 Place design

The transition (Figure 5) is just the reduction of all the inputs, sets transition activity as an output and if the transition is active and enable input is active, the increment output will be set to logical 1.

Mapping is just connecting the inputs and outputs according to the model.

Chooser is based on TRNG component called Figaro [9]. The output initializes the counter and the counter counts the number of shifts in LSFR that determines which transition will remain active.

Controller is the brain of the circuit, it has to keep up all the behavior of the PN. It controls the flow of tokens and drives the determination of which transition to choose.

3.1 PNML2VHDL

Modern companies try to achieve the control as simple as they can, so the SW implements the conversion without any need of knowledge of PNML and VHDL. Therefore the automatic parser according the Figure 7 was implemented.

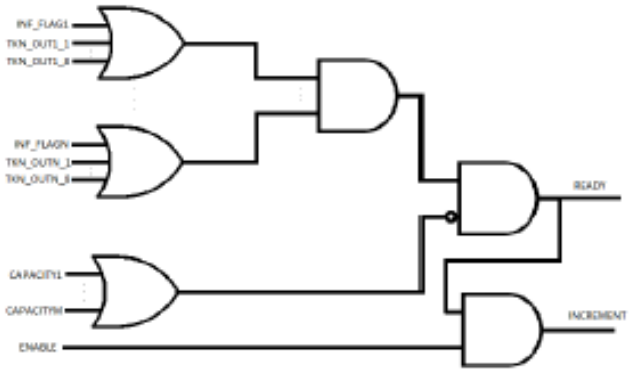

Figure 5 Transition design

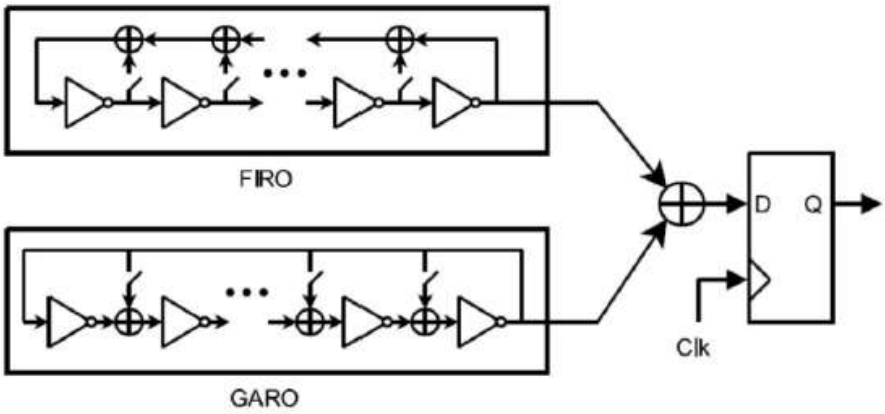

Figure 6 TRNG generator Figaro

5 Experiments

Several experiments were realized to prove the concept. The example of a simulation is done as test-beching the PN block and checking the flow of all transitions, and the result is in Figure 8.

7 Conclusions and Future Work

The previous steps can be applied using standard Petri nets, but our final aim is to construct a more complicated model (and a software tool) that will be applied to different types of Petri Nets (like coloured Petri Nets, Timed PNs) and not only the standard type, so our final model will consider particularly security and safety issues that the given system can experience. There are the following partial aims for our study: Study of methods used for certification safety and reliability parameters.

- Taking into account the appropriate metrics for quantification of different parameters (eg. the determination of resistance to attacks);
- Use of experiments performed by other DDD members and mutual interactions search (active involvement into the DDD research group);
- Design of algorithms and selection of suitable existing models or their modification to predict dependability and safety levels;

- Incorporation these algorithms into automated digital design methods to pre-select the ratio between size, reliability and safety or other requirements (low-power, hardware-software ratio and co-design, verification, testability).

Petri nets are very powerful and useful , and by adding them to proper tools we can get a great results that will help us in all work areas .Therefore, our resulted models will have more efficiency and less cost and of course they will be more secured and safe. But proper improvements of Petri Nets to be able verify such properties must be encapsulated and possible composition of other models and tools (Markov chains, UML diagrams) should be evolved.

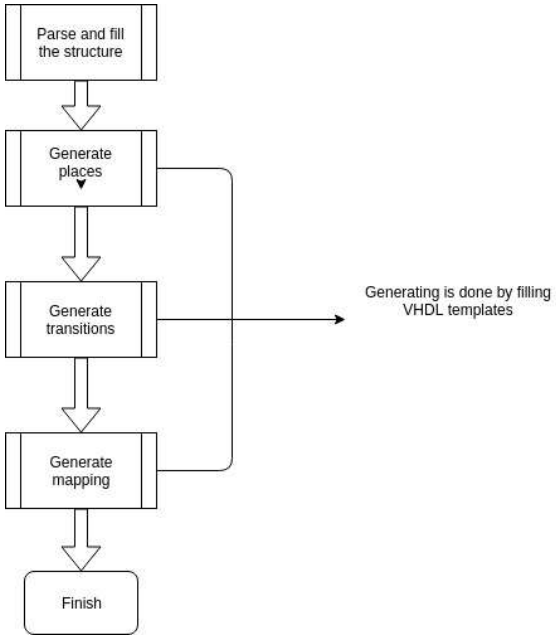

Figure 7 PNML to VHDL diagram

Acknowledgment

This paper is supported by Czech Technical University grant SGS17/213/OHK3/3T/18. I would like to thank doc. Ing. Hana Kubatova for her active guidance throughout my whole study.

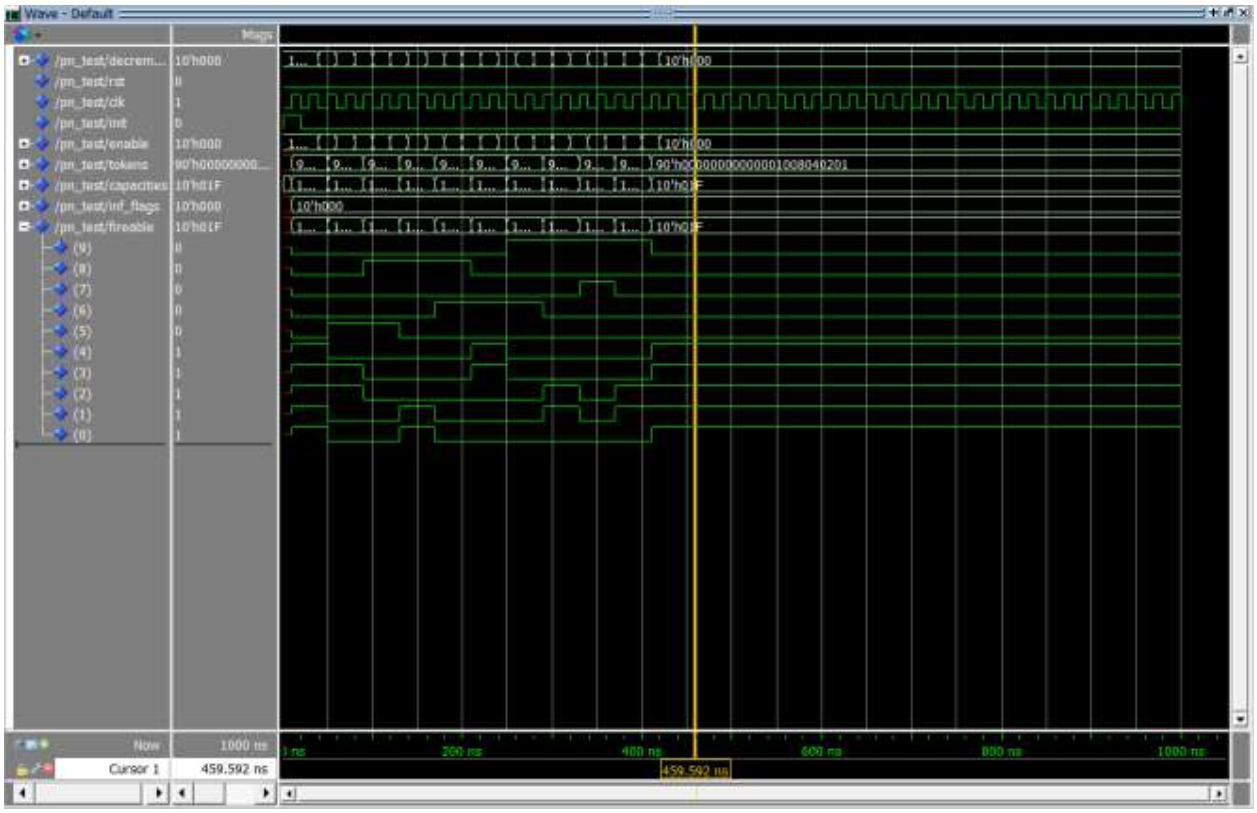

Figure 8 Simulation results

References

[1] Jakš, Z.:. Implementation of Petri net in FPGA. Master thesis. Prague: Czech Technical University in Prague, Faculty of information technology, 2018.

[2] Desrochers, A. and R. Ai-Jaar: Applications of Petri Nets in Manufacturing Systems: Modeling, Control AQ3 and Performance Analysis. IEEE Press, 1995.

[3] Cabasino M.P., Giua A., Seatzu C.: Introduction to Petri Nets. In: Seatzu C., Silva M., van Schuppen J. (eds) Control of Discrete-Event Systems. Lecture Notes in Control and Information Sciences, vol 433. Springer, London, 2013.

[4] M. Hack: The recursive equivalence of the reachability problem and the liveness problem for Petri nets and vector addition systems, FOCS, 1974, 156-164.

[5] P. Habermehl: On the complexity of the linear time mu-calculus for Petri nets, 18th International Conference on Application and Theory of Petri Nets, LNCS 1248, Springer-Verlag, Berlin, 1997, 102-116.

[6] R. Howell, P. Jancar and L. Rosier: Completeness results for single path Petri nets, Information and Computation 106 (1993), 253-265.

[7] Murata, T. 1989: Petri nets: Properties, analysis and applications. Proceedings of the IEEE 77(4): 541–580

[8] PNML standard: http://www.pnml.org/

[9] M. Dichtl, Jovan Dj, Golic: High-Speed True Random Number Generation with Logic Gates Only. HES 2007, LNCS 4727, pp. 45–62, 2007.

Author Index

Sponsors

EATON

Eaton is a power management company providing energy-efficient solutions that help our customers effectively manage electrical, hydraulic and mechanical power. A global technology leader, Eaton acquired Cooper Industries plc in November 2012. The 2012 revenue of the combined companies was $21.8 billion on a pro forma basis. Eaton has approximately 102,000 employees and sells products to customers in more than 175 countries.

STMicroelectronics

STMicroelectronics is a world leader in providing the semiconductor solutions that make a positive contribution to people's lives, today and into the future. ST is a global semiconductor company with net revenues of US$ 8.35 billion in 2017. Offering one of the industry's broadest product portfolios, ST serves customers across the spectrum of electronics applications with innovative semiconductor solutions for Smart Driving and the Internet of Things. By getting more from technology to get more from life, ST stands for life.augmented.

ASICentrum

ASICentrum, established in 1992 in Prague is a design center of EM Microelectronic and a competence center of ETA, belonging to the Swatch Group. EM Microelectronic is one of the most innovative IC providers. It developed and manufactured the smallest and the lowest power consuming Bluetooth chip on the market, the top performing optical sensors for optical office as well as gaming mice and it was the first to release the award-winning world-first dual-frequency NFC + RAIN RFID emlecho.

ESET

CZ.NIC

CZ.NIC, interest association of legal entities, was founded by leading providers of Internet services in 1998. The association currently has 114 members. The key activities of the association include operation of the domain name registry for the .CZ domain, operation of the CZ top-level domain and public education in the area of domain names. The association is now intensively working on development of the DNSSEC technology and mojeID service, extension and improvements of the domain administration system and support of new technologies and projects beneficial to the Internet infrastructure in the Czech Republic. CZ.NIC is a member of the EURid association, managing the European domain – EU, and other similarly oriented organizations (CENTR, ccNSO etc.).

CESNET

Czech Technical University in Prague

The conference has been sponsored by the CTU grant SVK 50/18/F8.

IEEE Student Branch at Czech Technical University in Prague

Partners

IEEE - Czechoslovakia Section, Computer Society