AFRL-AFOSR-VA-TR-2020-0130

# Monoidal computers, networks and strategic learning: Methods for Adaptive Defense in Cyber Security (MADCybS)

Dusko Pavlovic
UNIVERSITY OF HAWAII SYSTEMS HONOLULU

08/12/2020
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. REPORT DATE** *(DD-MM-YYYY)*
14-08-2020

**2. REPORT TYPE**
Final Performance

**3. DATES COVERED** *(From - To)*
15 Sep 2015 to 29 Feb 2020

**4. TITLE AND SUBTITLE**
Monoidal computers, networks and strategic learning: Methods for Adaptive Defense in Cyber Security (MADCybS)

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
FA9550-15-1-0263

**5c. PROGRAM ELEMENT NUMBER**
61102F

**6. AUTHOR(S)**
Dusko Pavlovic

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
UNIVERSITY OF HAWAII SYSTEMS HONOLULU
2530 DOLE STREET SAK D-200
HONOLULU, HI 96822-2309 US

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

**10. SPONSOR/MONITOR'S ACRONYM(S)**
AFRL/AFOSR RTA2

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
AFRL-AFOSR-VA-TR-2020-0130

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
A DISTRIBUTION UNLIMITED: PB Public Release

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**
We mapped the conceptual problem area and identified technical toolkits for the solutions. We made significant advances in a few technical tasks and small advances in several others. We defined the conceptual framework for recognizing and analyzing deception attacks and influence campaigns. An apparent paradigm shift in concept analysis emerged.

**15. SUBJECT TERMS**
Moving-Target Defense, Strategic Learning, Computability, Game Model, Monoidal Networks

| 16. SECURITY CLASSIFICATION OF: a. REPORT | b. ABSTRACT | c. THIS PAGE | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UU | | NGUYEN, TRISTAN |
| | | | | | **19b. TELEPHONE NUMBER** *(Include area code)* 703-696-7796 |

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

DISTRIBUTION A: Distribution approved for public release.

Final Report submitted to
Air Force Office of Scientific Research

# Monoidal Computers, Networks and Strategic Learning: Methods for Adaptive Defense in Cyber Security (MADCybS)

**Award No. FA9550-15-1-0263**

**Period of performance:** September 15, 2015 – February 29, 2020

**Principal Investigator:** Dusko Pavlovic (University of Hawaii)

# Contents



# 1 Work

## 1.1 Problem

This project was concerned with mathematical foundations of strategic reasoning in cyber security. The need for mathematically based strategic reasoning emerged with the Cold War, and was addressed by game theory, and by John von Neumann's strategic paradigm of *Mutual Assured Destruction (MAD)*. The strategic paradigm of cyber security is usually denoted by the phrase *Advanced Persistent Threat (APT)*. It does not yield to standard game theoretic tools. The goal of the present project was to analyze this gap, and the necessary mathematical updates. More precisely, we proposed to tackle the gap between

- game theoretic methods, based on the concept of equilibrium, and
- security methods, tools and protocols, based on cryptography and on computational complexity.

This gap is a result of two main differences:

a) *different roles of rules:*

   - game theory assumes that the rules of the games are efficiently enforced, that they are static, and that the players derive static preferences from them; whereas
   - security is the process where the participants seek to subvert or circumvent the enforced rules or protocols.

b) *different roles of computation:*

   - game theoretic methods are not stated in a computational form, and there is no uniform or canonical way to restrict them to a computational form; whereas

- security depends on modern cryptography, which fundamentally depends on the various computational assumptions, and cyber security is concerned with networks of computers.

The urgency of this work significantly increased during the period of performance, as the focus of the APT strategies shifted from malware-based penetrations into computer systems, to deceit-based influence of cyber-social groups and networks. Attackers' efforts progressed from technical exploits, to tactical coordination towards incremental strategic advances and buildup. The tasks of this project: to develop scientific models of security processes and the corresponding methods for strategic reasoning — shifted from a gap on the edge of the scene of security, to the center stage.

### 1.2 Summary of results

We mapped the conceptual problem area and identified the realm of technical toolkits for the solutions. It occupies large swaths of computational and mathematical semantics, computational linguistics, and social sciences. We made significant advances in a few technical tasks [1, 2, 3] and small advances in several others [4, 5, 6, 7]. An apparent paradigm shift in concept analysis emerged [8].

### 1.3 Novel insights

We proposed to develop a mathematical model of gaming capable to capture the strategic processes of *outsmarting*, that include phenomena such as *posturing* and *deceit*. Such phenomena have not been captured in the extant models of strategy design because they require taking into account the differences in players' logical and computational powers powers. In the final phase of the period of performance, an outline of a theory of deceit emerged, and gave us a glimpse of *scalable* methods to recognize and mitigate influence campaigns and the novel family of *level above* attacks.

The idea is that strategic learning in general, and outsmarting in particular, are implemented along the vertical axis of the *language stack*, which is better known in computer science as the *network stack*. The idea and the parallelism between the two, aligned through the *programming language stack*, is displayed in Fig. 1.3. Intuitively, a context can be construed as the semantical feature that allows a performer of a Turing Test to distinguish a human from a computer. A computer can be programmed to capture the meaning of words, and it can learn to recognize and generate correct sentences; but the *"threads of meaning"* that connect sentences into a conversation, or into a narrative, will remaining a problem. These "threads of meaning" are the context. E.g., if we take a novel, and remove half of its sentences, randomly selected, a human reader will usually perform much better than a computer in the task of guessing where the gaps are. Computers are capable of understanding and generating correct sentences; but at this time, only the humans understand and generate the contexts.

Just like a grammar determines the next word in a sentence as a probability distribution over a set of words, a context determines the next sentence in a text as a probability distribution over a set of sentences. The more unlikely choices of words are made, the more unexpected sentences are pronounced, the more information gets conveyed.

With the advent of network-based artificial intelligence, the idea of context acquires a new dimension. Computer networks nowadays process data at high semantical levels. For the web, this means that the

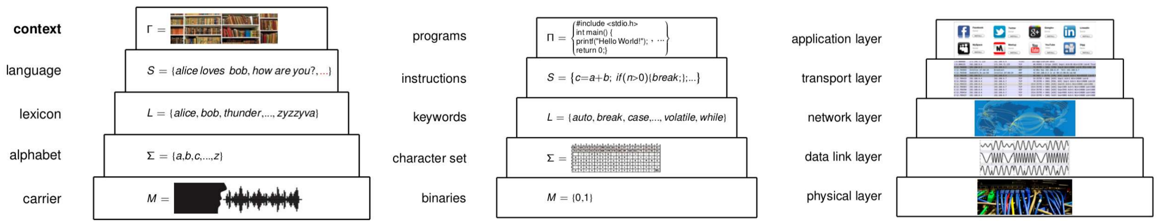


Figure 1: Semantical stacks

information flows not only on the level of *web contents*, i.e. through hypermedia, contained within the web pages, which are contained within the web sites, and so on; but the information also flows on the level of *web contexts*, which can be mined as *combinations* of web contents used within the same web service. For the internet, this means that the information flows not only on the level of packets and routing, but also on the level of *internet contexts*, which can be mined as frequent combinations of data flows that arise within the same services. The crucial insights that open an alley towards formalizing the idea of context for network interactions are:

1. network communication links are always layered:
    - a channel is always implemented as a data transmission layer over a physical transmission layer,
    - a language always consists of a set of words (a lexicon) built from a set of letters (the alphabet).
2. a previously realized network link can be used as the bottom layer for a link at the next network layer:
    - a physical link can be used to implement a data link, a data channel can be used to transmit encrypted signals, and thus carry a secure channel, or to transmit packets formatted for routing, and thus carry a transport layer;
    - a given set of letters can be used to build words, a given set of words can be used to build a sentences; a given set of sentences can be used to build contexts.

This *double articulation* of languages, where letters are grouped into words and words are grouped into sentences, has been put forward by the founders of structural linguistics (starting with Martinet and Hjelmslev) as the characterizing property of languages. The idea that the layering of language continues further up, as sentences are grouped into narratives, narratives into texts, and so on, has been studied in semiotics and philosophy, but no attempts have been made to formalize it. As computer networks spanned the cyberspace, and the social and computational interactions blended, the layered architecture of language been embodied in the layered network architecture. The four main layers which seem to occur in all networks and languages are displayed in Table 1. The leftmost column corresponds to the original concept of the internet architecture, of which the seven layer OSI network model is a later engineering extension. The other columns display the same idea where it was not explicitly engineered, but evolved through communication and use. The same architecture is, of course, substantially differently realized in the different areas, and there does not seem to exist a common terminology that aligns the corresponding layers and functionalities across the domains. But all cases seem usefully subsumed under a common mathematical model, based on actor networks. Their layered structure is displayed in the diagrammatic view of actor networks: an actor is viewed as a "state machine" where each "state" may contain another "state machine", where each "state" may again contain a

| *internet* | *speech* | *writing* | *programming* | *web* |
|---|---|---|---|---|
| links | voice, hearing | screen (paper) | binaries | internet |
| data | phonemes | letters | unicode | hypermedia |
| transport | morphemes | words | instructions | web pages |
| applications | sema | sentences | programs | web sites |
| SERVICES | NARRATIVES | TEXTS | COMPONENTS | SERVICES |

Table 1: Layered architecture of channels and sources

still lower level ”state machine”, etc. An example showing this view of a shared resource is displayed in Figure 2. An actor network is thus a multi-layer ”state machine”. Just like a finite state machine presents

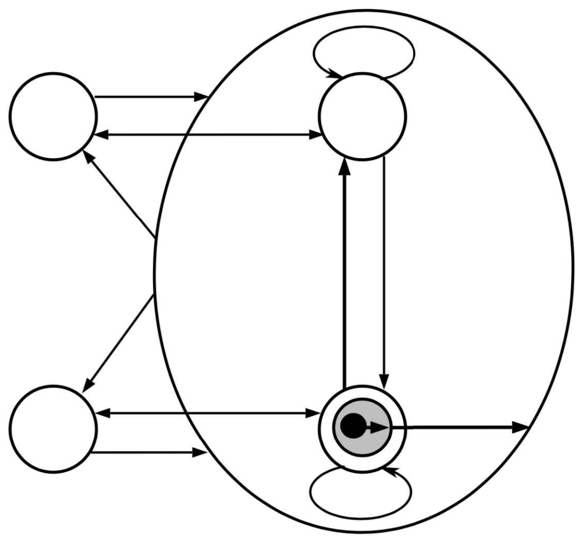

Figure 2: An actor network as a multi-layer ”state machine”

a language as a set of words, an actor network provides a view of language with its multiple articulations, or of the underlying multi-layer network of communication channels. A formal model of contexts spanned by the sentences of this language then emerges as the top layer of such actor network. Such contexts are displayed in the bottom row of Table 1. In each column, the bottom entry thus corresponds to the top layer of the displayed channel or source (since the layers are listed top-down, with the bottom layer in the first row).

### 1.4 Conclusion

Our investigations have demonstrated that the problem of strategic reasoning in cybersecurity stretches far, not only beyond the conceptual horizons of any single research project, but also beyond the reach of any of the existing research approaches and communities. It requires a security science, genuinely unified not only in name, but also in method. Cyberspace provides the attacker with a unified field of attack vectors, and requires a unified field of defense strategies from the defender. This unified field will have to be cultivated by a unified security science. A high-level map of this science was described in [3]. A more detailed picture is being drawn in [9].

# 2 Dissemination

## 2.1 Publications


[1] Toshiki Kataoka and Dusko Pavlovic. Towards Concept Analysis in Categories: Limit Inferior as Algebra, Limit Superior as Coalgebra. In Lawrence S. Moss and Pawel Sobocinski, editors, *Proceedings of CALCO 2016*, volume 35 of *LIPIcs*, pages 130–155, Dagstuhl, Germany, 2016. Leibniz-Zentrum für Informatik.

[2] Dusko Pavlovic and Peter-Michael Seidel. Quotients in monadic programming: Projective algebras are equivalent to coalgebras. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, June 20-23, 2017*, pages 1–12, 2017. arxiv:1701.07601.

[3] Jason Castiglione, Dusko Pavlovic, and Peter-Michael Seidel. Privacy protocols. In Joshua Guttman, editor, *CathFest: Proceedings of the Symposium in Honor of Catherine Meadows*, volume 11565 of *Lecture Notes in Computer Science*, pages 167–192. Springer, 2019.

[4] Vladimir Vovk and Dusko Pavlovic. Universal probability-free prediction. *Ann. Math. Artif. Intell.*, 81(1-2):47–70, 2017. arxiv.org:1603.04283.

[5] Dusko Pavlovic and Muzamil Yahia. Monoidal computer III: A coalgebraic view of computability and complexity. In Corina Cîrstea, editor, *Coalgebraic Methods in Computer Science (CMCS) 2018 — 14th IFIP WG 1.3 International Workshop, Revised Selected Papers*, volume 11202 of *Lecture Notes in Computer Science*, pages 167–189. Springer, 2018. arxiv:1704.04882.

[6] José-Luiz Fiadeiro, Ionut Tutu, Antonia Lopez, and Dusko Pavlovic. Logics for Actor Networks: A Case Study in Constrained Hybridization. *J. of Logical and Algebraic Methods in Programming*, 106:141 – 166, 2019.

[7] Roberto Bruni, Roberto Giacobazzi, Roberta Gori, Isabel Garcia-Contreras, and Dusko Pavlovic. Abstract Extensionality: On the properties of incomplete abstract interpretations. *Proc. ACM Program. Lang.*, 4(POPL):28:1–28:28, 2020.

[8] Dusko Pavlovic and Dominic J.D. Hughes. The nucleus: Mining concepts from adjunctions. arXiv:2004.07353.

[9] Dusko Pavlovic and Peter-Michael Seidel. *Security Science (SecSci) I: Ideas, logics and geometry of security, privacy, and trust*. Textbook in preparation. Fragments on asecolab.org. Current version (90 pp) available on request, 2021.

[10] Linda Briesemeister, Grit Denker, Karsten Martiny, Dusko Pavlovic, and Mark St. John. Policy creation for enterprise-level data sharing. In Abbas Moallem, editor, *Human-Computer Interaction for Cybersecurity, Privacy and Trust (HCI-CPT) 2019*, volume 11594 of *Lecture Notes in Computer Science*, pages 249–265. Springer, 2019.

[11] Dusko Pavlovic and Temra Pavlovic. Causality and deceit: Do androids watch action movies? *CoRR*, abs/1910.04383, 2019. to appear.

[12] Jason Castiglione and Dusko Pavlovic. Dynamic distributed secure storage against ransomware. *IEEE Transactions on Computational Social Systems*, 2019. Early Access in July 2019.

[13] Andrea Corradini, Tobias Heindel, Barbara König, Dennis Nolte, and Arend Rensink. Rewriting abstract structures: Materialization explained categorically. In *Foundations of Software Science and Computation Structures - 22nd International Conference, FOSSACS 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6-11, 2019, Proceedings*, pages 169–188, 2019.

[14] Benjamin Cabrera, Tobias Heindel, Reiko Heckel, and Barbara König. Updating probabilistic knowledge on condition/event nets using bayesian networks. In *29th International Conference on Concurrency Theory, CONCUR 2018, September 4-7, 2018, Beijing, China*, pages 27:1–27:17, 2018.

[15] Filippo Bonchi, Pierre Ganty, Roberto Giacobazzi, and Dusko Pavlovic. Sound up-to techniques and Complete abstract domains. In *33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 9-12, 2018*, pages 1–12, 2018. arxiv:1804.10507.

[16] Filippo Bonchi, Joshua Holland, Dusko Pavlovic, and Paweł Sobociński. Refinement for signal flow graphs. In Roland Meyer and Uwe Nestmann, editors, *Proceedings of CONCUR 2017*, volume 85 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[17] Depeng Li, Rui Zhang, Yingfei Dong, Fangjin Zhu, and Dusko Pavlovic. A multisecret value access control framework for airliner in multinational air traffic management. *IEEE Internet of Things Journal*, 4(6):1853–1867, 2017.

[18] Filippo Bonchi, Dusko Pavlovic, and Paweł Sobociński. Functorial Semantics for Relational Theories. Technical report, ASECOLab, November 2017. arxiv:1711.08699.

[19] José-Luiz Fiadeiro, Ionut Tutu, Antonia Lopez, and Dusko Pavlovic. Logics for Actor Networks: A Case Study in Constrained Hybridization. In Alexandre Madeira and Mário Benevides, editors, *Proceedings of DaLi 2017 — Dynamic Logic and Applications*, volume 10669 of *Lecture Notes in Computer Science*, pages 98–114. Springer, 2017.

[20] Dusko Pavlovic and Bertfried Fauser. Smooth coalgebra: testing vector analysis. *Mathematical Structures in Computer Science*, 27(7):1195–1235, 2017.

[21] Dusko Pavlovic and Peter-Michael Seidel. (Modular) effect algebras are equivalent to (Frobenius) antispecial algebras. In Ross Duncan and Chris Heunen, editors, Proceedings 13th International Conference on *Quantum Physics and Logic,* Glasgow, Scotland, 6-10 June 2016, volume 236 of *Electronic Proceedings in Theoretical Computer Science*, pages 145–160. Open Publishing Association, 2017.

[22] Vladimir Vovk and Dusko Pavlovic. Universal probability-free conformal prediction. In Alexander Gammerman et al, editor, *Proceedings of 5th International Symposium on Conformal and Probabilistic Prediction with Applications (COPA)*, volume 9653 of *Lecture Notes in Computer Science*, pages 40–47. Springer, 2016.

[23] Dusko Pavlovic. *MonCom: Basic Concepts of Computation in Diagrams*. Textbook in preparation. Fragments on asecolab.org. Current version (220 pp) available on request., 2020.

## 2.2 Researchers and students

**David Basin** (ETH Zürich, Professor, Director of the Information Security Group, Head of Department of Computer Science): collaborator in ongoing work on privacy protocol analysis;

**Filippo Bonchi** (U. of Pisa, Associate Professor): coauthor of [15, 16, 18];

**Roberto Giacobazzi** (U. of Verona, Professor): coauthor of [15];

**Whitfield Diffie** ("Father of Modern Cryptography"): joined the project in September 2017, working on applying the project results towards *absolute one-way functions*;

**Tobias Heindel** (U. of Hawaii, postdoc): coauthor of [13] and [14];

**Dominic Hughes** (Apple Inc.): coauthor of [8];

**Peter-Michael Seidel** (U. of Hawaii, Associate Professor): coauthor of [2];

**Pawel Sobocinski** (U. of Tartu, Professor): coauthor of [16] and [18];

**Vladimir Vovk** (Royal Holloway, Professor): coauthor of [4];

**Muzamil Yahia** (U. of Hawaii, graduate student): coauthor of [5];

## 2.3 Outreach

The PI coorganized

- ExtInt workshop:
  `http : //shonan.nii.ac.jp/seminar/115/`.
  The post-proceedings of the workshop, based on the project research, will be published by Springer.
- CathyFest:
  `https : //link.springer.com/content/pdf/bfm%3A978 − 3 − 030 − 19052 − 1%2F1.pdf`