

Kibernetinio saugumo reikalavimų užtikrinimas kritinės infrastruktūros sistemose naudojant formalaus verifikavimo ir dirbtinio intelekto metodus

Daniel Daukševič

Vilniaus Universitetas
Matematikos ir Informatikos Fakultetas
daniel.dauksevic@mif.stud.vu.lt

Ataskaitinė Informatikos krypties doktorantų konferencija
2026 m. spalio 8 d.



Doktorantūra

Tema

Kibernetinio saugumo reikalavimų užtikrinimas kritinės infrastruktūros sistemose naudojant formalaus verifikavimo ir dirbtinio intelekto metodus

Mokslo kryptis

Informatika N009

Vadovas

prof. dr. Linas Laibinis

Studijų laikotarpis

2023 m. lapkričio mėn. 1 d. - 2027 m. spalio mėn. 31 d. (I metai)

Tyrimas

Objektas

Kibernetinio saugumo reikalavimai KI sistemose

Tikslas

Sukurti metodiką, leidžiančią formalizuoti kibernetinio saugumo reikalavimus modeliuojant KI tinklus

Planuojami rezultatai

KI sistemų kibernetinio saugumo vertinimo ir reikalavimų modeliavimo metodikos

Uždaviniai

- Iširti formalius metodus, tinkamus KI sistemų kibernetinio saugumo reikalavimų formaliajam verifikavimui
- Iširti DI metodus, taikytinus verifikavimo parametrų identifikavimui ar optimalumo kriterijų derinimui

Turinys

- 1 Studijos
- 2 Moksliniai tyrimai
- 3 Kito pusmečio planas
- 4 Kito pusmečio planas

1 Studijos

2 Moksliniai tyrimai

3 Kito pusmečio planas

4 Kito pusmečio planas

Studijų planas

Studijų metai	Egzaminai ¹	
	Planas	Ivykdyta
I (2023/2024)	1	1
II (2024/2025)	2	2
III (2025/2026)	1	
IV (2026/2027)		
Iš viso:	4	3

Lentelė 1. Studijų planas ir jo vykdymo suvestinė

Studijuojami dalykai

Egzaminai		
Planas	Įvykdyta	Būklė
Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika 2024 m. II ketvirtis	Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika, 2024-06-25	Išlaikyta
Fundamentalieji informatikos ir informatikos inžinerijos metodai 2025 m. I ketvirtis	2025 m. I ketvirtis	Išlaikyta
Automatizuoti verifikavimo ir validavimo metodai 2025 m. II ketvirtis	2025 m. II ketvirtis	Išlaikyta
Gilieji neuroniniai tinklai 2026 m. II ketvirtis	2026 m. II ketvirtis	-

Lentelė 3. Mokslinių tyrimų ir disertacijos rengimo etapai

Gilieji neuroniniai tinklai

Learning-Driven Statistical Model Checking

- Duomenų generavimas
 - Tinklo modeliavimas (UPPAAL)
 - Sintetinių duomenų generavimas
 - Žymėjimas (UPPAAL SMC)
- Neuroninio tinklo architektūros kūrimas
 - Regresijos/klasifikacijos modelis (PyTorch/TensorFlow)
 - Sustiprintas mokymasis (RL)
- Eksperimentinis vertinimas

1 Studijos

2 Moksliniai tyrimai

3 Kito pusmečio planas

4 Kito pusmečio planas

Planuojamas mokslinių tyrimų planas

PLANUOJAMAS MOKSLINIŲ TYRIMŲ PUBLIKAVIMAS

Preliminari mokslinės publikacijos tema, numatomas mokslo leidinys		Data
1. Tyrimo "Requirements of Cybersecurity of Critical Infrastructure Systems" rezultatų publikavimas tarptautiniame recenzuojamoje mokslo žurnale arba konferencijoje (Pvz. ECCWS, SAFECOMP)		2025 m. II ketvirtis
2. Tyrimo "Methods for modelling cyber security requirements for critical infrastructure systems" rezultatų publikavimas tarptautiniame recenzuojamoje mokslo žurnale arba konferencijoje (Pvz. ECCWS, SAFECOMP)		2026 m. III ketvirtis
3. Tyrimo "Formal verification of the requirements ensuring the security of critical infrastructure systems" rezultatų publikavimas žurnaluose su citavimo rodikliu (WoS IF) (Pvz. Journal of cybersecurity, Journal of IEEE, IT IEEE Transactions on Dependable and Secure Computing Professional, IEEE Transactions on Reliability, Formal Aspects of Computing)		2026 m. IV ketvirtis
4. Tyrimo "Ensuring and improving the cyber security of critical infrastructure systems using formal verification and artificial intelligence" rezultatų publikavimas žurnaluose su citavimo rodikliu (WoS IF) (Pvz. Journal of cybersecurity, Journal of IEEE, IT IEEE Transactions on Dependable and Secure Computing Professional, IEEE Transactions on Reliability, Formal Aspects of Computing)		2027 m. II ketvirtis

Mokslinių tyrimų suvestinė

<u>Studijų metai</u>	<u>Dalyvavimas konferencijose</u>				<u>Publikacijos</u>					
	<u>Tarptautinėse²</u>		<u>Nacionalinėse³</u>		<u>Su citav. rodikliais⁴</u>			<u>Be citav. rodiklio⁵</u>		
	<u>Planas</u>	<u>Ivykdyta</u>	<u>Planas</u>	<u>Ivykdyta</u>	<u>Planas</u>	<u>Ivykdyta⁶</u>	<u>Būklė⁷</u>	<u>Planas</u>	<u>Ivykdyta⁶</u>	<u>Būklė⁷</u>
I (2023/2024)	0	0			0			0		
II (2024/2025)	1	0			0			1	0	
III (2025/2026)	1	2			1		Ruošiama	1	2	
IV (2026/2027)	0				1			0		
<u>Iš viso:</u>	2	2			2	0		0	2	

Lentelė 2. Doktorantūros studijų planas ir jo vykdymo suvestinė

Dalyvavimas konferencijose

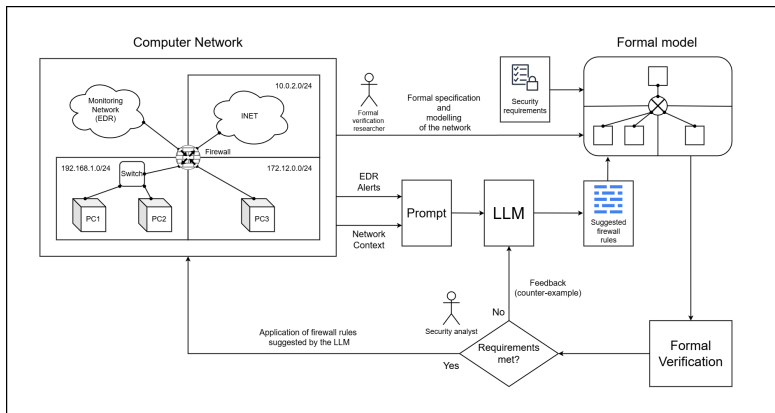
- 14-oji jaunųjų mokslininkų konferencija „Fizinių ir technologijos mokslų tarpdalykiniai tyrimai“
- 70th International Conference for Students of Physical and Natural Sciences "Open Readings 2026"

Studijų pasiekimai

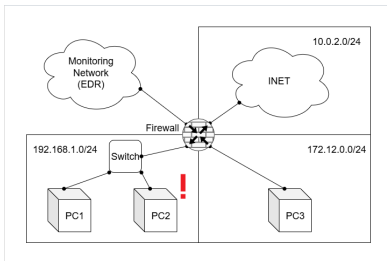
<u>Dalyvavimas tarptautinėse konferencijose</u>	
Nr.	<u>Aprašas</u>
1	D. Daukševič, R. Magán-Carrión and L. Laibinis, "Towards Verified Security: Formal Methods for LLM-Based Defenses," ICC 2026 - IEEE International Conference on Communications, Glasgow, United Kingdom, 2026.
2	D. Daukševič, L. Laibinis, "Formal Verification of Healthcare Computer Network Architectures Using Alloy and TLA+", ABZ 2026 - Rigorous State-Based Methods, Tokyo, Japan, 2026.
3	D. Daukševič, "Cybersecurity Requirements Assurance in Critical Infrastructure Systems through the Application of Formal Verification and Artificial Intelligence-Based Methods", Doctoral Symposium, FM 2026 - "International Symposium on Formal Methods", Tokyo, Japan, 2026.

Lentelė 5. Doktorantūros studijų pasiekimai

I. Towards Verified Security: Formal Methods for LLM-Based Defenses



Towards Verified Security: Formal Methods for LLM-Based Defenses



```
iptables -A FORWARD -s 192.168.1.0/24 -d  
→ 192.168.1.0/24 -j DROP # This means every  
→ incoming or outgoing traffic in subnet A  
→ will be dropped.
```

Towards Verified Security: Formal Methods for LLM-Based Defenses

Action to forward or block the packet

```
RoutePacket  $\triangleq$ 
   $\wedge$  QueuedIDs  $\neq \{\}$ 
   $\wedge$  LET  $min\_id \triangleq$  CHOOSE  $id \in$  QueuedIDs :  $\forall other \in$  QueuedIDs :  $id \leq other$ 
    packet  $\triangleq$  CHOOSE  $p \in$  queued_packets :  $p.id = min\_id$ 
  IN
   $\wedge$  queued_packets' = queued_packets \ {packet}
   $\wedge$  IF Forwarding between subnets allowed, sender and receiver not blocked
     $\wedge$  (packet.srcSubnet, packet.dstSubnet)  $\in$  ALLOWED_ROUTES
     $\wedge$  packet.src  $\notin$  BLOCKED_SENDERS
     $\wedge$  packet.dst  $\notin$  BLOCKED_RECEIVERS
  THEN Forward packet
     $\wedge$  received_packets' = [received_packets EXCEPT ![packet.dst] = @  $\cup$  {packet}]
   $\wedge$  UNCHANGED blocked_packets
  ELSE Block packet
     $\wedge$  blocked_packets' = blocked_packets  $\cup$  {packet}
   $\wedge$  UNCHANGED received_packets
   $\wedge$  UNCHANGED {packets, packet_num}
```

All received devices have allowed route, sender and receiver

```
PacketReceivedRouteAllowed  $\triangleq$ 
   $\forall ip \in$  DOMAIN received_packets :
     $\forall packet \in$  received_packets[ip] :
       $\wedge$  (packet.srcSubnet, packet.dstSubnet)  $\in$  ALLOWED_ROUTES
       $\wedge$  packet.src  $\notin$  BLOCKED_SENDERS
       $\wedge$  packet.dst  $\notin$  BLOCKED_RECEIVERS
```

All received devices has secure sender and receiver

```
PacketReceivedDevicesSecure  $\triangleq$ 
   $\forall ip \in$  DOMAIN received_packets :
     $\forall packet \in$  received_packets[ip] :
       $\wedge$  SecureDevice(packet.src)
       $\wedge$  SecureDevice(packet.dst)
```

Towards Verified Security: Formal Methods for LLM-Based Defenses

The screenshot displays a formal verification tool interface. On the left, a sidebar shows 'Model Checking Results' with a link to 'See query execution results'. The main area is divided into two panes. The left pane shows a table of states with columns: Time, Dis., State, District St., Queue, Module, Action, Location, State Found, and District Status. The right pane shows a 'Model' section with a 'State' field and a 'Next State' field. Below this, there is a 'Error Trace' section showing a sequence of states and actions. The bottom of the interface has a 'Progress Output' section.

Time	Dis.	State	District St.	Queue	Module	Action	Location	State Found	District Status
0000000	0	000	000	000	000	000	000	000	000
0000001	0	1	1	1	000	000	000	000	000
0000002	0	1	1	1	000	000	000	000	000
0000003	0	1	1	1	000	000	000	000	000
0000004	0	1	1	1	000	000	000	000	000
0000005	0	1	1	1	000	000	000	000	000
0000006	0	1	1	1	000	000	000	000	000
0000007	0	1	1	1	000	000	000	000	000
0000008	0	1	1	1	000	000	000	000	000
0000009	0	1	1	1	000	000	000	000	000
0000010	0	1	1	1	000	000	000	000	000
0000011	0	1	1	1	000	000	000	000	000
0000012	0	1	1	1	000	000	000	000	000
0000013	0	1	1	1	000	000	000	000	000
0000014	0	1	1	1	000	000	000	000	000
0000015	0	1	1	1	000	000	000	000	000
0000016	0	1	1	1	000	000	000	000	000
0000017	0	1	1	1	000	000	000	000	000
0000018	0	1	1	1	000	000	000	000	000
0000019	0	1	1	1	000	000	000	000	000
0000020	0	1	1	1	000	000	000	000	000
0000021	0	1	1	1	000	000	000	000	000
0000022	0	1	1	1	000	000	000	000	000
0000023	0	1	1	1	000	000	000	000	000
0000024	0	1	1	1	000	000	000	000	000
0000025	0	1	1	1	000	000	000	000	000
0000026	0	1	1	1	000	000	000	000	000
0000027	0	1	1	1	000	000	000	000	000
0000028	0	1	1	1	000	000	000	000	000
0000029	0	1	1	1	000	000	000	000	000
0000030	0	1	1	1	000	000	000	000	000
0000031	0	1	1	1	000	000	000	000	000
0000032	0	1	1	1	000	000	000	000	000
0000033	0	1	1	1	000	000	000	000	000
0000034	0	1	1	1	000	000	000	000	000
0000035	0	1	1	1	000	000	000	000	000
0000036	0	1	1	1	000	000	000	000	000
0000037	0	1	1	1	000	000	000	000	000
0000038	0	1	1	1	000	000	000	000	000
0000039	0	1	1	1	000	000	000	000	000
0000040	0	1	1	1	000	000	000	000	000
0000041	0	1	1	1	000	000	000	000	000
0000042	0	1	1	1	000	000	000	000	000
0000043	0	1	1	1	000	000	000	000	000
0000044	0	1	1	1	000	000	000	000	000
0000045	0	1	1	1	000	000	000	000	000
0000046	0	1	1	1	000	000	000	000	000
0000047	0	1	1	1	000	000	000	000	000
0000048	0	1	1	1	000	000	000	000	000
0000049	0	1	1	1	000	000	000	000	000
0000050	0	1	1	1	000	000	000	000	000
0000051	0	1	1	1	000	000	000	000	000
0000052	0	1	1	1	000	000	000	000	000
0000053	0	1	1	1	000	000	000	000	000
0000054	0	1	1	1	000	000	000	000	000
0000055	0	1	1	1	000	000	000	000	000
0000056	0	1	1	1	000	000	000	000	000
0000057	0	1	1	1	000	000	000	000	000
0000058	0	1	1	1	000	000	000	000	000
0000059	0	1	1	1	000	000	000	000	000
0000060	0	1	1	1	000	000	000	000	000
0000061	0	1	1	1	000	000	000	000	000
0000062	0	1	1	1	000	000	000	000	000
0000063	0	1	1	1	000	000	000	000	000
0000064	0	1	1	1	000	000	000	000	000
0000065	0	1	1	1	000	000	000	000	000
0000066	0	1	1	1	000	000	000	000	000
0000067	0	1	1	1	000	000	000	000	000
0000068	0	1	1	1	000	000	000	000	000
0000069	0	1	1	1	000	000	000	000	000
0000070	0	1	1	1	000	000	000	000	000
0000071	0	1	1	1	000	000	000	000	000
0000072	0	1	1	1	000	000	000	000	000
0000073	0	1	1	1	000	000	000	000	000
0000074	0	1	1	1	000	000	000	000	000
0000075	0	1	1	1	000	000	000	000	000
0000076	0	1	1	1	000	000	000	000	000
0000077	0	1	1	1	000	000	000	000	000
0000078	0	1	1	1	000	000	000	000	000
0000079	0	1	1	1	000	000	000	000	000
0000080	0	1	1	1	000	000	000	000	000
0000081	0	1	1	1	000	000	000	000	000
0000082	0	1	1	1	000	000	000	000	000
0000083	0	1	1	1	000	000	000	000	000
0000084	0	1	1	1	000	000	000	000	000
0000085	0	1	1	1	000	000	000	000	000
0000086	0	1	1	1	000	000	000	000	000
0000087	0	1	1	1	000	000	000	000	000
0000088	0	1	1	1	000	000	000	000	000
0000089	0	1	1	1	000	000	000	000	000
0000090	0	1	1	1	000	000	000	000	000
0000091	0	1	1	1	000	000	000	000	000
0000092	0	1	1	1	000	000	000	000	000
0000093	0	1	1	1	000	000	000	000	000
0000094	0	1	1	1	000	000	000	000	000
0000095	0	1	1	1	000	000	000	000	000
0000096	0	1	1	1	000	000	000	000	000
0000097	0	1	1	1	000	000	000	000	000
0000098	0	1	1	1	000	000	000	000	000
0000099	0	1	1	1	000	000	000	000	000
0000100	0	1	1	1	000	000	000	000	000

```
iptables -A FORWARD -s 192.168.1.2 -d
192.168.1.2 -j DROP # This means all
incoming or outgoing traffic of PC2 will be
dropped.
```

The 43rd IEEE International Conference on Communications (IEEE ICC 2026)

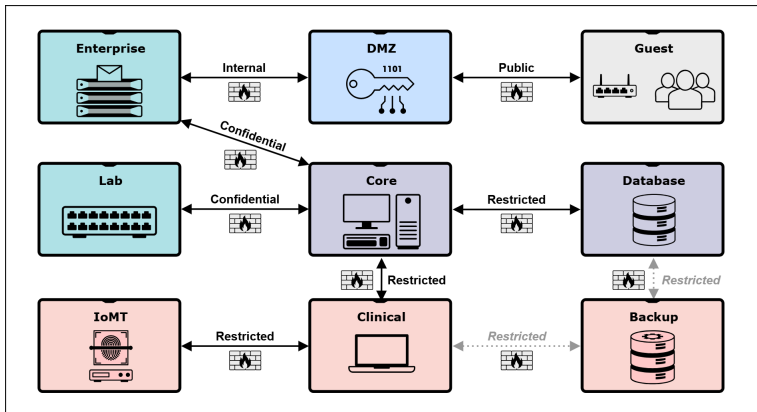
- Gegužės 24–28, Glazgas, Škotija
- Ruošiamas straipsnis publikavimui žurnale su citavimo rodikliu



Tyrimo tęsinys

- Kotrpavyzdžio formalizavimas LLM pateikimui
- Kitų formalaus verifikavimo įrankių taikymas (pvz. SMT Solvers)
- Kitų kibernetinio saugumo reikalavimų modeliavimas (pvz. confidentiality)
- Struktūriškai sudėtingesnių tinklų modeliavimas

II. Formal Verification of Healthcare Network Architectures



Static Property Analysis in Alloy

SP3. Architectural Integrity: *Every hop in every path shall connect zones whose trust levels are equal or differ by at most one level in total ordering.*

SP3 $\triangleq \forall p \in P : \forall i \in \{1, \dots, |p|\} :$

$\text{trustLevel}(\text{src}(p[i])) = \text{Next}(\text{trustLevel}(\text{dst}(p[i])))$

$\vee \text{trustLevel}(\text{dst}(p[i])) = \text{Next}(\text{trustLevel}(\text{src}(p[i])))$

$\vee \text{trustLevel}(\text{src}(p[i])) = \text{trustLevel}(\text{dst}(p[i]))$

Dynamic Property Analysis in TLA⁺

DP1. Dynamic Reconfiguration: *Upon detection of compromise in any zone with trust level AZ or CZ, the network shall eventually transition to Reconfiguring mode.*

$$\text{DP1} \triangleq \Box((\exists z \in Z : \text{zoneCompromised}(z) \wedge \text{trustLevel}(z) \in \{AZ, CZ\}) \\ \rightarrow \Diamond(\text{networkState} = \text{Reconfiguring}))$$

The 12th International Conference on Rigorous State Based Methods (ABZ 2026)

- Gegužės 18-22, Tokijas, Japonija
- Pradedami darbai sekančiam straipsniui, skirtam publikacijai žurnale su citavimo rodikliu



The 27th International Symposium on Formal Methods (FM 2026) Doctoral Symposium



- 1 Studijos
- 2 Moksliniai tyrimai
- 3 Kito pusmečio planas**
- 4 Kito pusmečio planas

Disertacijos rengimo etapai

MOKSLINIŲ TYRIMŲ IR DISERTACIJOS RENGIMO ETAPAI

Darbo pavadinimas	Atlikimo terminai	
	Nuo	Iki
1. Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):		
1.1. Disertacijos tyrimo objekto detalizavimas	2024 m. I ketvirtis	2024 m. I ketvirtis
1.2. Mokslinės literatūros ir publikacijų analizė	2024 m. I ketvirtis	2025 m. IV ketvirtis
1.3. Kitų publikacijų analizė (standartai, reikalavimai ir pan.)	2024 m. I ketvirtis	2025 m. IV ketvirtis
2. Mokslinio tyrimo vykdymas:		
2.1. Tyrimo metodikos sudarymas:		
2.1.1. Tyrimo metodikos iškeltiems uždaviniams spręsti parinkimas	2024 m. I ketvirtis	2024 m. III ketvirtis
2.1.2. Teorinio ir empirinio tyrimų (pagal pasirinktą metodiką) planų sudarymas	2024 m. IV ketvirtis	2025 m. II ketvirtis
2.2. Teorinis tyrimas:		
2.2.1. Kritinės infrastruktūros sistemų saugumo pažeidžiamų lygmenų įtakos tyrimas bei taksonomijos pažeidžiamumo paviršiai aprašyti sudarymas	2024 m. III ketvirtis	2026 m. I ketvirtis
2.2.2. Kritinės infrastruktūros sistemų saugumo užtikrinimo reikalavimų modeliavimo apjungiant formalaus verifikavimo ir dirbtinio intelekto metodų tyrimas	2025 m. III ketvirtis	2026 m. III ketvirtis
2.3. Empirinis tyrimas:		
2.3.1. Kritinės infrastruktūros sistemų saugumo užtikrinimo reikalavimų formalaus verifikavimo tyrimas (parinktų pavyzdžio sistemų pagrindu)	2026 m. I ketvirtis	2027 m. I ketvirtis
2.3.2. Kritinės infrastruktūros sistemų saugumo užtikrinimo sprendimų ir taikomų metodų optimizavimas ir tobulinimas taikant formalaus verifikavimo ir dirbtinio intelekto įrankius (parinktų pavyzdžio sistemų pagrindu)	2026 m. III ketvirtis	2027 m. II ketvirtis
2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas:		
2.4.1. Gautų duomenų analizė	2026 m. II ketvirtis	2027 m. I ketvirtis
2.4.2. Rezultatų apibendrinimas ir išvadų parengimas	2027 m. I ketvirtis	2027 m. II ketvirtis
3. Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:		
3.1. Tyrimo metodikos sudarymas	2024 m. I ketvirtis	2025 m. II ketvirtis
3.2. Teorinės disertacijos dalies parengimas	2024 m. III ketvirtis	2026 m. III ketvirtis
3.3. Eksperimentinės disertacijos dalies parengimas	2026 m. I ketvirtis	2027 m. II ketvirtis
3.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas	2026 m. I ketvirtis	2027 m. II ketvirtis
4. Daktaro disertacijos parengimas ir svarstymas padalinyje:	2027 m. I ketvirtis	2027 m. III ketvirtis
5. Daktaro disertacijos gynimas:	2027 m. III ketvirtis	2027 m. IV ketvirtis

- 1 Studijos
- 2 Moksliniai tyrimai
- 3 Kito pusmečio planas
- 4 Kito pusmečio planas

Kito pusmečio darbo planas

- Pateikti pirmą straipsnį publikavimui žurnale su citavimo rodikliu;
- Pradėti ruošti sekantį straipsnį publikavimui žurnale su citavimo rodikliu;
- Pradėti gautų rezultatų analizės apibendrinimą ir išvadų parengimą