# A REAL-TIME DEEP PACKET INSPECTION INTRUSION DETECTION SYSTEM FOR SOFTWARE DEFINED 5G NETWORKS

Dr. Razvan Bocu

Department of Mathematics and Computer Science

Transilvania University of Brasov, Romania


DEEPSEC

# ABSTRACT

**Abstract** The philosophy that founds the world of the Internet of Things apparently becomes essential for the projected permanently connected world. The 5G data networks are supposed to dramatically improve the actual 4G networks' real world significance, which makes them fundamental for the next generation networks of IoT devices. The academic and industrial effort to improve the 5G technological standards considers various routes. Thus, this paper presents the state-of-the-art concerning the development of the standards that model the 5G networks. It values the authors' experience that was gathered during the implementation of the Vodafone Romania 5G networked services. It puts this acquired experience in context by reviewing the relevant similar work, the relevant technologies, and it describes the research directions and difficulties that will probably influence the design and implementation of large 5G data networks. Consequently, the paper presents a machine learning-based real time intrusion detection system, which has been effectively tested in the context of a 5G data network. The intelligent intrusion detection system considers the creation of software defined networks, and it uses artificial intelligence based models. It is able to detect unknown intrusions through the usage of machine learning-based software components. The system has been assessed and the results prove that it achieves superior performance with a lower overhead in comparison to similar approaches, which allows it to be deployed on real-time 5G networks.

# AGENDA

- Technical Requirements
- Relevant Technologies
- Presentation of the Intrusion Detection System
- Performance Assessment
- Conclusions


DEEPSEC

# TECHNICAL REQUIREMENTS

- High throughput data channels – the deployment of the smart applications requires data links that offer transfer speeds of, at least, 25 Mbps, which are intended to support the high definition data containers, the virtual reality (VR) or augmented reality (AR) applications
- Networks that are scalable and structurally flexible – this is determined by the consideration of the mechanism of network functions virtualization (NFV) in order to build the required fronthaul data networks
- Very low latency – the 5G IoT networks are intended to support smart applications that should send and receive real-time data, which require communication channels with latencies that are no greater than 5 milliseconds


DEEPSEC

# TECHNICAL REQUIREMENTS

- Reliability and resilience – the existence of sensibly more small network cells in a 5G data network involves that the handover should be conducted in an efficient way, while the network coverage is kept at the optimal levels
- Data privacy and security – the deployment of applications that process highly senstive data, such as personal health data, implies that the proper mechanisms should be designed and implemented in order to prevent any illegitimate access attempt
- Long battery lifetime – the mobility is a central concept in the realm of 5G data networks, and consequently the energy efficiency should be considered


DEEPSEC

# TECHNICAL REQUIREMENTS

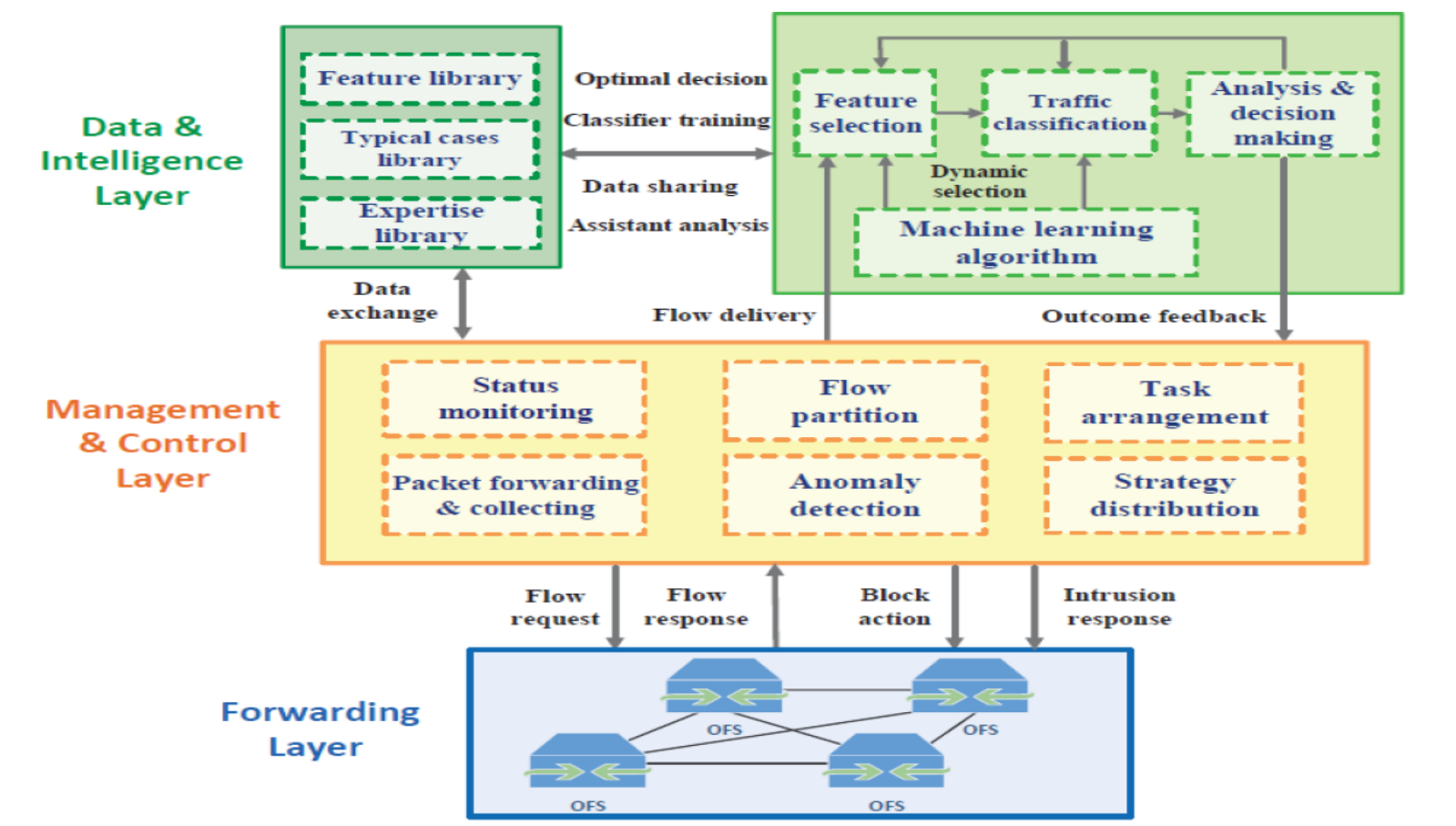
- Connection density – the 5G data networks are expected to offer concurrent reliable access for a large number of devices, which implies that proper design and implementation decisions are made
- Mobility – this technical requirement complements the necessity to offer proper conditions for the deployment of many devices, which require reliable mobile intercommunication data links


DEEPSEC

# RELEVANT TECHNOLOGIES

The virtualized wireless network function (VWNF) is a fundamental process in the realm of the 5G networks design and implementation. It has been effectively used in order to deploy the core of the mentioned 5G network. This process allows for the logical specification of a self-sufficient 5G network by using the network functions virtualization (NFV) on the proper hardware infrastructures. It is immediate to note that this process is interesting from a research and theoretical perspective. Furthermore, it allows for the specialized 5G networks to be deployed on specific infrastructures, such as cloud infrastructures, or telecommunications service providers networks [8]. We have effectively used this mechanism in order to implement specialized networked services on the 5G data network of the respective telecommunications service provider. We have observed that the virtualized networked environment offers the required logical flexibility and scalability, which allowed us to efficiently deploy the real-time intrusion detection system

# THE INTRUSION DETECTION SYSTEM



# THE INTRUSION DETECTION SYSTEM

- The architecture of the system consists of three layers: the **data traffic forwarding** layer, the **data management and control** layer, and the **machine learning-based data analysis** layer. The data forwarding layer is responsible for the data traffic monitoring and capturing. It can collect and send the suspect data streams to the control layer, and it also blocks the malicious data traffic according to the instructions of the controller. The data management and control layer identifies the suspicious data patterns, and detects anomalies using the analyzed intercepted data. It also takes proper protection measures according to the decisions made by the data analysis layer, and it consequently instructs the data forwarding layer.


DEEPSEC

# THE INTRUSION DETECTION SYSTEM

- The data forwarding layer provides the data management and control layer with real-time network status information through the real-time collection of suspect data patterns. Furthermore, intrusions are immediately blocked by dropping the malicious packects under the supervision of the other system layers.
- The packet collection and data flow partitioning layer provides a more global view of the entire 5G network. The status monitoring module supervises the data network status and continuously analyzes the data packets that it receives in order to analyze them. The data management and control layer processes and parses the received data traffic. Furthermore, it creates relevant clusters of data packets and generates a data fingerprint, which keeps track of the following logical network parameters: the source IP address, the destination IP address, the source port, the destination port, the session duration, and the considered network protocol.


DEEPSEC

# THE INTRUSION DETECTION SYSTEM

- The data fingerprints are used in order to define and label different data flow records, which represent specific network connections and activities. The packet collection and inspection is performed continuously. The data collection and inspection time interval is optimized in order to avoid possible undesirable delays concerning the real-time data analysis process. The anomaly detection considers some basic flow statistics, which are used to roughly recognize abnormal behaviors and potential anomalies. The particular intrusion detection system’s module applies an entropy-based analysis, which is based on the Shannon’s theory in order to detect the distribution variations of the analysed data packet samples.


DEEPSEC

# THE INTRUSION DETECTION SYSTEM

- The entropy of a random variable x is computed considering the following formula:

$$H(x) = -\sum_{i=1}^{n} p(x_i) log(p(x_i))$$


DEEPSEC

# THE INTRUSION DETECTION SYSTEM

- Here, $p(x_i)$ designates the probability for $x$ to take the value $x_i$ considering all the already detected values. The equation considers four fundamental parameters: the source IP address, the source port, the destination IP address, and the destination port. The values of these parameters are gathered by the real-time traffic analysis component of the system. Thus, considering a particular moment in time, the continuously updated value that the entropy function $H(x)$ provides helps to detect possible malicious data traffic patterns. Considering that $E$ stands for the mean entropy, and $S$ represents the corresponding standard deviation, a possible suspect pattern involves that the value of $H(X)$ is outside the interval $[(E-S),(E+S)]$. Consequently, the suspect data packets are sent over to the proactive data analysis layer for supplementary analysis.


DEEPSEC

# THE INTRUSION DETECTION SYSTEM

- The feature selection component is designed in order to construct and update the features set, which is specific to the detected malicious data patterns. This component is capable to process large amounts of data in a real-time fashion, while removing the data features that are irrelevant to the machine learning core of the system’s proactive data analysis layer. Consequently, the data is partitioned into relevant categories, so that malicious data traffic patterns are clearly separated from the benign traffic patterns.
- The feature selection component is designed in order to construct and update the features set, which is specific to the detected malicious data patterns. This component is capable to process large amounts of data in a real-time fashion, while removing the data features that are irrelevant to the machine learning core of the system’s proactive data analysis layer. Consequently, the data is partitioned into relevant categories, so that malicious data traffic patterns are clearly separated from the benign traffic patterns.


DEEPSEC

# PERFORMANCE ASSESSMENT

- The performance analysis considers the data that was effectively gathered during the real-time intrusion detection process on the provider’s 5G data network. The dataset that was considered for the performance assessment contains 32,000,000 analyzed network connections. Each individual connection entity consists of 39 features that are separated into three categories. Thus, the system considers network connections-based features, content-based features, and data traffic-based features. Furthermore, each data traffic entity is marked either as a normal traffic entity, or as a suspicious traffic entity. The latter ones are grouped into four distinct categories: remote to local, probe, user to root, and denial of service.


DEEPSEC

# PERFORMANCE ASSESSMENT

- The performance assessment considers the following metrics: precision (P), reliability (R), tradeoff (T), accuracy (A), and the false positives rate (FP). The precision is defined as the percentage of valid malicious data traffic predictions relative to the total number of predictions that the intrusion detection system makes. The reliability is calculated as the total number of accurately determined intrusion attempts relative to the total number of intrusions. Furthermore, the tradeoff represents a hybrid performance metric between the precision and the reliability, which has the role to provide a better accuracy of the data classification through the following formula: $T = 2 \triangleleft ((1 \triangleleft P) + (1 \triangleleft R))$


DEEPSEC

# PERFORMANCE ASSESSMENT

- The accuracy is a ratio that is determined by the sum of the number of legitimate packets and malicious packets properly detected at the numerator, while the denominator is the sum of the accurately detected legitimate and malicious packets plus the incorrectly detected legitimate and malicious packets. Moreover, the false positives rate is determined by the number of legitimate packets that are incorrectly classified over the sum between properly classified legitimate packets and incorrectly classified legitimate packets.


DEEPSEC

# PERFORMANCE ASSESSMENT

| Data size | P | R | T | A | FP |
|---|---|---|---|---|---|
| 20% | 97.21% | 96.87% | 94.24% | 94.13% | 0.86% |
| 40% | 97.02% | 96.53% | 94.05% | 93.92% | 1.03% |
| 60% | 96.13% | 95.89% | 93.68% | 93.39% | 0.93% |
| 80% | 96.04% | 95.81% | 93.48% | 93.18% | 0.91% |
| 100% | 95.47% | 95.12% | 93.04% | 92.07% | 1.06% |


DEEPSEC

# CONCLUSIONS

- The system scales well with the size of the analyzed data set. Furthermore, the system is able to accurately determine the malicious traffic patterns, while reducing to the minimum the incidence of the false positives. The practical behaviour of the system is especially important in the case of commercial 5G data networks, which transport and process a large number of data transfer sessions that have to be analyzed in a proactive manner.
- The 5G data networks already support relevant real-world applications, and they have the potential to become the backbone of the future always connected human society. Consequently, there are rather difficult design, implementation and deployment problems, which concern all aspects of the 5G networks. Among them, the timely detection of any illegitimate access attempt is essential, especially in the context of a commercial data network. Therefore, this paper presents the state-of-the-art concerning the research that has been made on this very important topic. Furthermore, a real-time intrusion detection system, which is based on the utilization of machine learning techniques, is described.


DEEPSEC

# CONCLUSIONS

- The performance of the system has been tested using real-world data, which has been obtained through the real-time monitoring of the 5G data traffic on the network of a significant Romanian telecommunications services provider. This assessment demonstrates that it is possible to design a software system that blocks most of the illegitimate traffic, which occurs on a high-traffic 5G data network, in a real-time fashion. Moreover, the various existing contributions, which are relevant to the approached topic, are presented in a constructive analytical manner, while the problems that have to be addressed are analyzed, and possible solutions are suggested for their resolution.


DEEPSEC

# CONTACT INFORMATION


- Dr. Razvan Bocu
- Department of Mathematics and Computer Science, Transilvania University of Brasov, Romania
- Email address: razvan.bocu@unitbv.ro



DEEPSEC

# Thank you for your attention!


DEEPSEC