Resilient NFV Service Chains under Energy-Aware Attacks: A Bilevel Optimization Approach

Mohammad Ali Raayatpanah*, Jocelyne Elias†, Fabio Martignon‡, Andrea Pimpinella‡, Michaël Poss§
* *Mathematical Sciences and Computer, Kharazmi University, Tehran, Iran*
† *Department of Computer Science and Engineering, University of Bologna, Italy*
‡ *Department of Management, Information and Production Engineering, University of Bergamo, Italy*
§ *LIRMM, University of Montpellier, CNRS, France*

***Abstract*—We investigate the problem of resilient and energy-aware Virtual Network Function (VNF) placement and routing in softwarized networks under the threat of targeted cyberattacks. We model the system as a bilevel interdiction game, where a malicious attacker strategically disrupts servers within a fixed resource budget, while a network provider reacts by minimizing energy consumption through optimized VNF deployment and flow routing. The lower-level problem includes capacity constraints, service function chaining, and a server energy model accounting for idle and load-dependent consumption. Attack-induced load shifts are captured via additive energy penalties on compromised nodes. To solve this inherently difficult bilevel integer program, we develop a single-level reformulation via interdiction cuts and propose a cutting-plane algorithm to explore the attacker's strategy space efficiently. Numerical experiments show the effectiveness of the approach in quantifying trade-offs between resilience and energy efficiency, supporting trustworthy and adaptive NFV deployment in critical infrastructures.**

***Index Terms*—Bilevel optimization, Virtual Network Functions, Cyber-physical resilience, Energy-aware networking, Interdiction modeling.**

I. Introduction

Recent advances in networking architectures, particularly through Network Function Virtualization (NFV) and Software-Defined Networking (SDN), have enabled more flexible and scalable network management. Together, they support Service Function Chains (SFCs), i.e., ordered sequences of virtual functions that network traffic must traverse to allow the provisioning and enforcement of specific services or policies (e.g., firewalls, load balancers or intrusion detection systems). In addition to optimizing service delivery, network operators must address security threats that target infrastructure vulnerabilities. Beyond operational risks, attack-induced load perturbations can lead to significant energy inefficiencies, resulting in higher costs and potential service outages due to power surges or budget exhaustion [2]. To mitigate these risks, operators often adopt energy-aware resource management strategies that not only improve energy efficiency but also enhance the network's resilience and robustness to security threats.

In this work, we investigate the problem of resilient and energy-aware Virtual Network Functions (VNFs) placement and routing under the threat of targeted cyberattacks [5]. We model the system as a bi-level interdiction game, where a malicious attacker strategically disrupts servers within a fixed resource budget, while a network provider reacts by minimizing total energy consumption through optimized VNF deployment and flow routing. The lower-level problem considers service function chaining while incorporating both link and node capacity constraints, alongside an energy consumption model that reflects idle and load-dependent usage. Energy penalties are added to compromised nodes to represent the impact of attack-induced load variations. To tackle this challenging bilevel integer program, we reformulate it as a single-level model using interdiction cuts, and propose a cutting-plane algorithm that systematically explores and refines the feasible set of interdiction strategies. Numerical results highlight the model's capability to assess the trade-offs between energy efficiency and cyber resilience in softwarized environments, and demonstrate the effectiveness of our approach in performing intelligent orchestration of network services under adversarial conditions, thus contributing to the development of resilient and energy-aware infrastructures for smart computing.

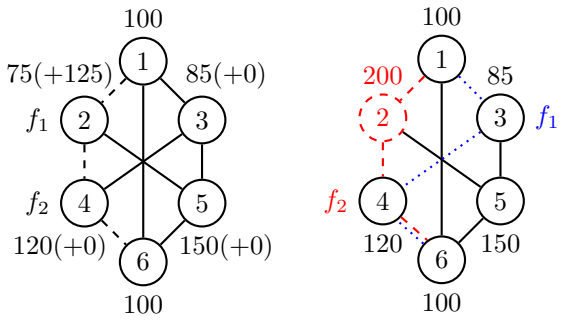
II. Illustrative Scenario

Figure 1 illustrates a fully connected network of six nodes linked via bidirectional links. In this scenario, node 1 issues a request for a SFC terminating at node 6. The SFC requires two sequential VNFs, f_1 and f_2, with f_1 preceding f_2. Each node may host a VNF only if its energy consumption remains below a predefined threshold (e.g., 150 units). Under normal conditions (Figure 1a), the traffic is routed from node 1 to node 2 for the execution of f_1, then to node 4 for f_2, yielding a total energy cost of 395 units. Figure 1b shows a compromised scenario where an attacker with limited interdiction budget targets node 2. The attack more than doubles the node's energy usage, making it ineligible to host f_1. A revised configuration preemptively places f_1 on node 3, effectively mitigating the impact of the attack. This adjustment increases energy consumption by just 2.5% under normal conditions. Importantly, f_2 remains on node 4, showing that only a partial reconfiguration is needed to maintain service continuity.

III. Mathematical Formulation

We first formalize the lower-level (follower) problem, which captures the VNF placement and routing with energy consumption minimization. We later extend it to the full bilevel interdiction model. Let $G = (V, A)$ be a (bi-)directed graph, where nodes $v \in V$ represent servers with core capacity c_v^{node}, and links $uv \in A$ have bandwidth capacity c_{uv}^{link} and cost η_{uv}. A set of VNFs F is available, each requiring a fraction δ_f of a CPU core and able to serve up to θ_f traffic units.

Requests $k \in K$ are defined by a source o^k, destination t^k, bandwidth demand b^k, and a service function chain (SFC) F^k,



(a) Normal conditions. Path cost: 395.

(b) Attack on Node 2. Dash path cost: 520. Dot path cost: 405.

Fig. 1: Example scenario, before (left) and after (right) interdiction of node 2. When interdicted, energy consumption of node 2 peaks at 200 units, 1.6 times the normal consumption.

an ordered list of required VNFs. Each request is routed along a single path.

Let us first define our problem decision variables: w_v is a binary server activation variable. z^k_{uv} is a binary flow routing variable, for request k. Integer variable γ^f_v denotes the number of VNF f installed at node v; y^k_{fv} specifies whether function f of request k is installed at node v or not. x^k_{fv} equals 1 if f of request k is installed at or before node v, and 0 otherwise.

a) Routing constraint: Flow conservation at each node enforces single-path routing:

$$\sum_{vu\in\delta^+(v)} z^k_{vu} - \sum_{uv\in\delta^-(v)} z^k_{uv} = \begin{cases} 1 & \text{if } v = o^k \\ -1 & \text{if } v = t^k \\ 0 & \text{otherwise.} \end{cases}$$

b) Link capacity: Total flow on each link must respect its bandwidth limit:

$$\sum_{k\in K} b^k z^k_{uv} \le c^{link}_{uv}, \quad \forall uv \in A.$$

c) VNF installation: Each VNF in the SFC of a request must be deployed exactly once:

$$\sum_{v\in V} y^k_{fv} = 1, \quad \forall f \in F^k,\ k \in K.$$

d) Node capacity: VNF instances on node v must not exceed its CPU capacity. κ_v is the total CPU consumed on v:

$$\kappa_v = \sum_{f\in F} \delta_f \gamma^f_v \le c^{node}_v w_v, \forall v \qquad \sum_{k\in K} y^k_{fv} b^k \le \theta_f \gamma^f_v, \quad \forall f, v.$$

e) SFC ordering: For each request, VNFs must be visited in the specified order:

$$y^k_{fv} \le x^k_{fv}, \quad x^k_{gv} \le x^k_{fv}, \quad (z^k_{uv} - 1) + (x^k_{fv} - x^k_{fu}) \le y^k_{fv}.$$

Energy Consumption/Cyberattacks in Network Servers: Operational energy costs in networks are driven by both static (idle) and dynamic (load-dependent) server consumption [6]. We adopt the following model to express total energy usage of server v as:

$$e_v(Tcap_v, Ccap_v) = (e^{\max}_v - e^{\text{idle}}_v)\frac{Ccap_v}{Tcap_v} + e^{\text{idle}}_v$$

where e^{idle}_v and $e^{\max}_v$ are idle and peak energy levels, respectively. $Tcap_v$ represents the server's total capacity, while $Ccap_v$ is the currently consumed capacity.

Cyberattack Impact: Malicious attacks (e.g., DoS, spoofing) can impair servers, increasing load and energy usage [8]. We model this using a constant additive factor γ_v, applied when node v is under attack. The adjusted energy function becomes:

$$e_v(Tcap_v, Ccap_v) = (e^{max}_v + \gamma_v - e^{idle}_v)\frac{Ccap_v}{Tcap_v} + e^{idle}_v. \quad (1)$$

Attack Budget and Interdiction Model: Attackers operate under a budget B, allocating resources p_v to attack nodes. Interdiction is modeled with binary variables $h_v, \in \{0, 1\}$, indicating if a node is under cyberattack. These affect node availability and capacity, and are governed by an *interdiction budget*: $\sum_{v\in V} p_v h_v \le B, \quad \forall v \in V$.

Bilevel Interdiction Game (IG): We now formalize the bilevel problem, where the attacker (upper level) maximizes disruption via targeted interdictions, and the provider (lower level) minimizes energy-aware VNF placement and routing costs.

$$z^* = \max_{h\in\mathcal{H}} \min_{(w,z,y,x,\kappa,\beta)\in\mathcal{X}} \mathcal{F}(w, \kappa)$$

where

$$\mathcal{F}(w, \kappa) = \sum_{v\in V} \left[e^{idle}_v w_v + (e^{max}_v + \gamma_v h_v - e^{idle}_v)\frac{\kappa_v}{c^{node}_v} \right]$$

s.t. *Routing, Capacity, Service Chaining, and Interdiction Budget constraints.*

The set $\mathcal{X}$ under which we are minimizing $\mathcal{F}(w, \kappa)$ represents the feasible solution space of the lower level/follower problem.

This IG model captures the interplay between strategic attacks and resilient VNF placement, with energy considerations integrated into both decision layers. The inner optimization, an integer program, poses significant computational challenges [1].

IV. Cutting-Plane based Single-Level Reformulation

We reformulate the bilevel interdiction problem as a single-level model using interdiction cuts [3, 7]. The follower's value function under a given interdiction $h \in H$ is:

$$\mathcal{I}(h) = \min_{(w,z,y,x,\kappa,\beta)\in\mathcal{X}} \mathcal{F}(w, \kappa) \quad (3)$$

We replace the follower's problem with constraints derived from optimal recourse responses, resulting in:

$$z^* = \max \eta \quad (4a)$$

$$\text{s.t. } \eta \le \mathcal{F}(w, \kappa) \quad \forall (w, z, y, x, \kappa, \beta) \in \mathcal{X} \quad (4b)$$

$$\sum_{v\in V} p_v h_v \le B. \quad (4c)$$

To manage the exponential number of cuts in (4b), a **cutting-plane method** (summarized in Algorithm 1) iteratively adds only violated constraints. We first solve the recourse problem without interdiction to get $\mathcal{I}(0)$, enforcing $\eta \ge \mathcal{I}(0)$.

An upper bound is: $z^* \le \sum_{v\in V} \left[e^{idle}_v + e^{max}_v + \gamma_v \right]$.

At each iteration, we solve the relaxed problem to get (η^*, h^*), then solve (3) for h^*. If a violated cut is found (i.e., $\mathcal{I}(h^*) < \eta^*$), we add: $\eta \le \mathcal{F}(w^*, \kappa^*)$. (5)

Initial Cut Generation procedure (ICG)

To initialize the cutting-plane algorithm, we generate an initial set of interdiction cuts based on the starting solution w^0 and budget B. Specifically, for each active node in the initial solution, we simulate an attack configuration that is both feasible and strategically impactful. We iteratively allocate B to nodes with high marginal energy impact per unit of cost, greedily selecting the node v that maximizes γ_v/p_v, until B is exhausted. Each simulated attack scenario leads to a re-evaluated network configuration, from which we extract an upper bound on the system's energy consumption. These bounds are then added as initial interdiction cuts of the form (5). These cuts restrict the feasible region of the upper-level problem by incorporating worst-case energy values under plausible interdiction strategies.

Algorithm 1 Cutting-Plane Method for Solving the IG Problem

Input: Relaxed master problem (4)
Add bounds: $\eta \leq \sum_v [e_v^{\text{idle}} + e_v^{\max} + \gamma_v]$
Generate initial cuts as explained in the ICG procedure
while violated cut exists **do**
    Solve relaxed problem $\rightarrow (\eta^*, h^*)$
    Solve (3) for h^* to get $\mathcal{I}(h^*)$
    if $\mathcal{I}(h^*) < \eta^*$ **then**
        Add violated cut as in the ICG procedure
    else
        Terminate: Optimal solution found
    end if
end while

V. Experimental Analysis

In this Section we test the effectiveness of the proposed optimization model on the *Abilene* topology (from the SNDLib database), consisting of 12 nodes and 30 bi-directional links.

1) Experiments Setup: We consider a network with three types of nodes, each having different computational and storage capabilities [4]: standard off-the-shelf servers (5 nodes), smart NICs (4 nodes), and PISA switches (3 nodes). The network demands correspond to four applications with different usage probability and required SFCs [4]: video streaming, web services, VoIP, and online gaming. We set B as the maximum number of nodes that can be interdicted, varying it within the range $[3, 6]$, and let γ_v be proportional to $e_v^{\max}$ by a factor α, varied within $[0.5, 1]$.

2) Performance Metrics: We consider as baseline metric the cost $z_{\text{no\_int}}$ of the optimal solution obtained when i) the attacker's strategy is not considered by the model and ii) no nodes are interdicted. The cost of such solution updates to $z_{\text{naive\_int}}$ if one or more network nodes get interdicted, according to the attacker's budget B and to the energy penalty α induced by the attack. Finally, we define z_{bilevel} as the cost obtained after applying the proposed bi-level optimization program.

3) Numerical Results: Table I reports the results obtained considering $|K| = [2, 4, 8]$ service requests. We also report the relative ratios $R_{\text{naive\_int}}$ and R_{bilevel} between $z_{\text{naive\_int}}$ and z_{bilevel} with respect to $z_{\text{no\_int}}$, respectively. We observe that interdiction attacks lead to significant cost increases when SFC demands are allocated without accounting for the attacker's strategy. In particular, $z_{\text{naive\_int}}$ is at least 63% greater than $z_{\text{no\_int}}$ in all scenarios, with such overhead doubling the baseline cost in the worst case scenario ($|K| = 2, \alpha = 1$). Remarkably, using the proposed framework reduces the overhead costs induced by attacks up to 83% ($|K| = 2, \alpha = 1$), at an average 40% higher costs with respect to non-interdicted network scenarios. Finally, larger $|K|$ leads the attacker to fully utilize its interdiction budget B, impacting costs significantly for $|K| = 8$.

TABLE I: Impact of $|K|$, B and α on solution costs.

$\lvert K\rvert$	B	α	$z_{\text{no\_int}}$	$z_{\text{naive\_int}}$	z_{bilevel}	$\text{R}_{\text{naive\_int}}(\%)$	$\text{R}_{\text{bilevel}}(\%)$
2	3	0.5	75	150	88	100	17
	3	1	75	225	100	200	33
	6	0.5	75	150	88	100	17
	6	1	75	225	100	200	33
4	3	0.5	120	195	170	63	42
	3	1	120	270	220	125	83
	6	0.5	120	195	170	63	42
	6	1	120	270	220	125	83
8	3	0.5	752	1278	940	70	25
	3	1	752	1803	1143	140	52
	6	0.5	752	1353	1015	80	35
	6	1	752	1952	1278	159	70

Acknowledgments

This work is supported by the University of Montpellier's Advanced Knowledge Institute on Transitions (MAK'IT) visiting scientist program, by projects *SERICS* (PE00000014) under the NRRP MUR program, funded by the EU-NGEU, and PRIN *NEWTON* (2022ZA8T22).

VI. Conclusions

This work addressed the problem of energy-aware VNF placement and routing under targeted cyberattacks by formulating a bi-level interdiction model. Computational experiments demonstrate the effectiveness of the proposed approach in anticipating and mitigating the impact of attacks. Compared to naive strategies, which do not consider the attacker's behavior, our method significantly reduces (on average by 59%) the energy consumption overhead induced by interdictions in the considered scenarios. These results highlight the importance of accounting for adversarial actions in the design of resilient and energy-efficient network infrastructures.

References

[1] F. Furini, I. Ljubić, P. San Segundo, and Y. Zhao. A branch-and-cut algorithm for the edge interdiction clique problem. *European Journal of Operational Research*, 294(1):54–69, 2021.

[2] H. Huang, W. Lin, J. Lin, and K. Li. Power management optimization for data centers: A power supply perspective. *IEEE Transactions on Sustainable Computing*, pages 1–20, 2025.

[3] M. Leitner, I. Ljubić, M. Monaci, M. Sinnl, and K. Tanınmış. An exact method for binary fortification games. *European Journal of Operational Research*, 307(3):1026–1039, 2023.

[4] R. Lin, L. He, S. Luo, and M. Zukerman. Energy-Aware Service Function Chaining Embedding in NFV Networks. *IEEE Transactions on Services Computing*, 16(2):1158–1171, 2023.

[5] Q. Lv, J. Zhu, F. Zhou, and Z. Zhu. Network Planning with Bilevel optimization to Address Attacks to Physical Infrastructure of SDN. In *ICC 2020*, pages 1–6, 2020.

[6] L. Niccolini, G. Iannaccone, S. Ratnasamy, J. Chandrashekar, and L. Rizzo. Building a power-proportional software router. In *USENIX ATC*, pages 89–100, Boston, Massachusetts, 2012.

[7] J. C. Smith and Y. Song. A survey of network interdiction models and algorithms. *European Journal of Operational Research*, 283(3):797–811, 2020.

[8] K. Yan, X. Liu, Y. Lu, and F. Qin. A cyber-physical power system risk assessment model against cyberattacks. *IEEE Systems Journal*, 17(2):2018–2028, 2023.