UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

ISLab Intelligent Systems Laboratory

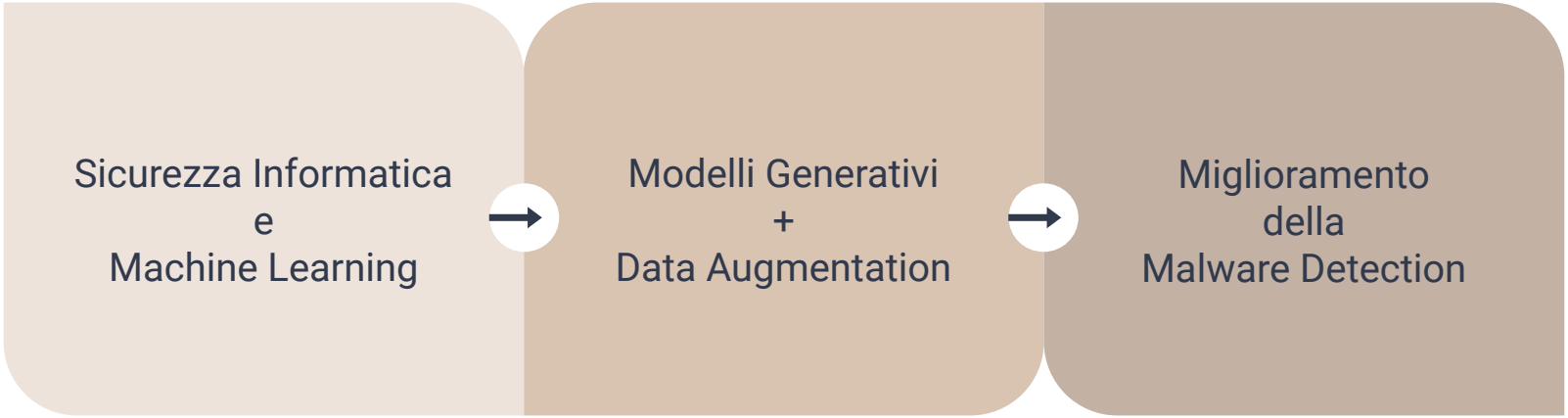
dib

Modelli Generativi di Deep Learning per la Cybersecurity

Prof. Donato Impedovo
Dott. Vincenzo Dentamaro
Dott. Stefano Galantucci

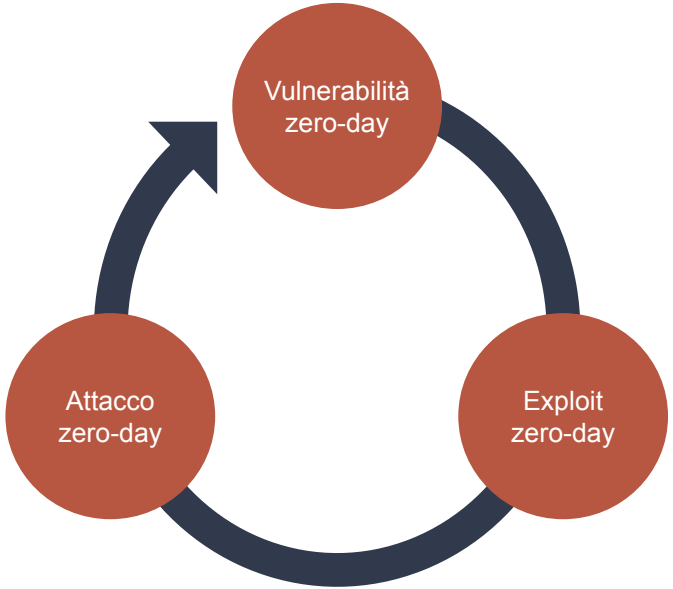
Gian Marco Ninno

Introduzione

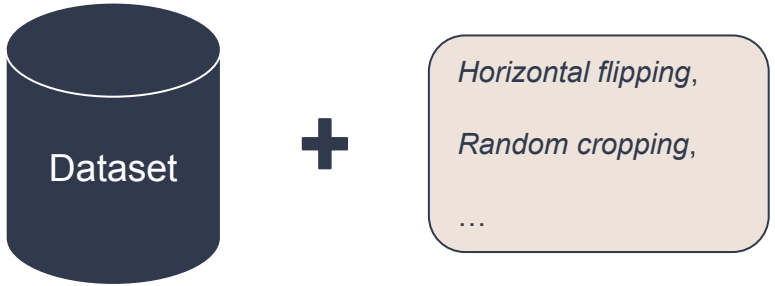


Problemi

Malware zero-day

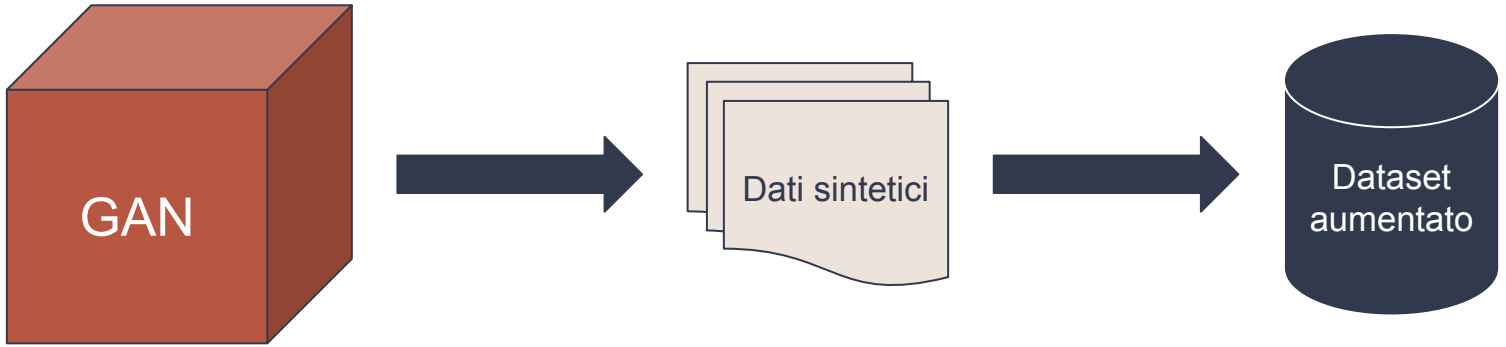


Data Augmentation tradizionale



Proposta

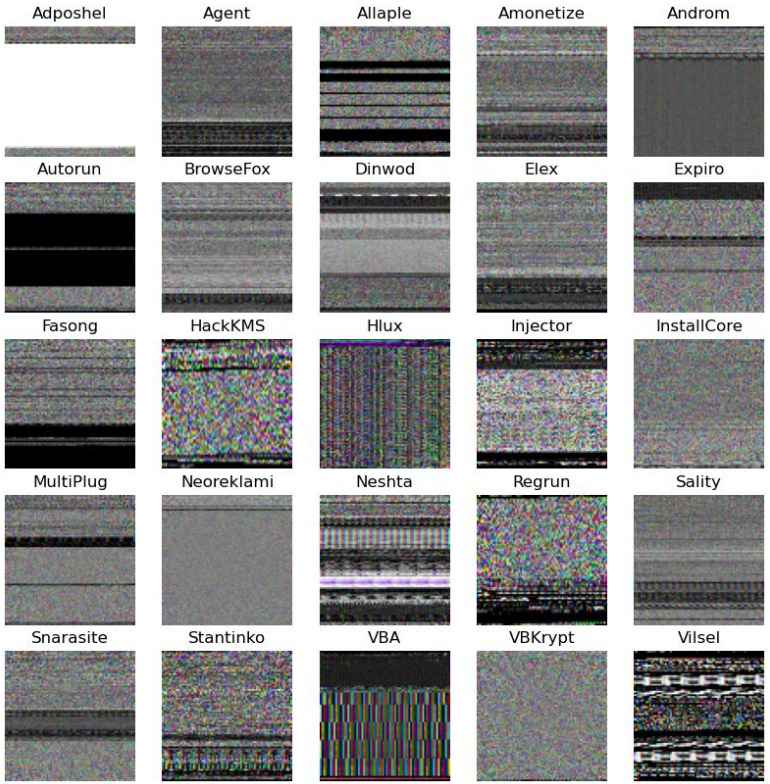
Data Augmentation con dati sintetici



Dataset MaleVis

Il dataset MaleVis (*Malware evaluation with Vision*) è formato da immagini RGB rappresentanti file PE (*Portable Executable*) malevoli del sistema operativo Microsoft Windows.

Gli esempi di questo dataset sono stati scelti da malware emersi nel 2017-2018.



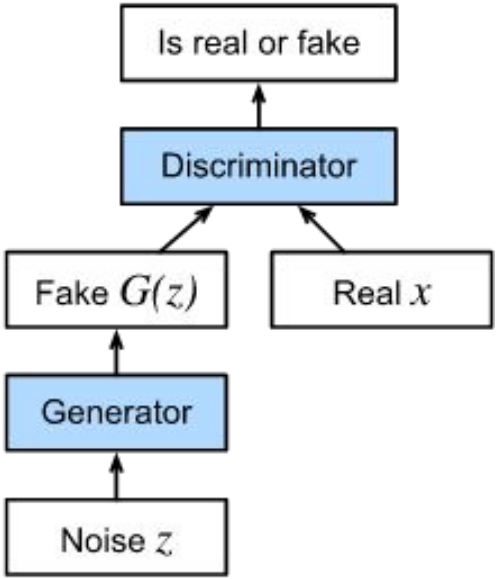
Generative Adversarial Network (GAN)

Il *framework* GAN è costituito da due reti:

- Generatore
- Discriminatore

Generatore e Discriminatore fanno un gioco a somma zero, in cui il progresso di uno porta all'indebolimento dell'altro.

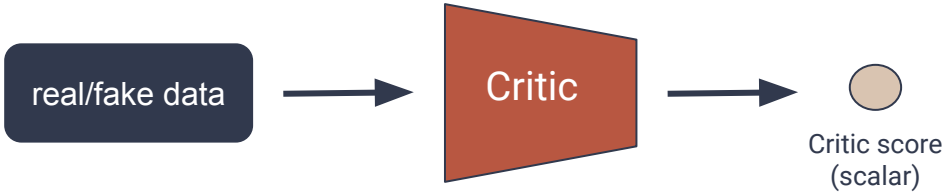
L'obiettivo è raggiungere un equilibrio tra le due reti.



Problemi comuni nelle GAN

- *Vanishing Gradient*
- *Mode Collapse*
- *Non convergenza*

Wasserstein Conditional GAN with Gradient Penalty (WCGAN-GP)



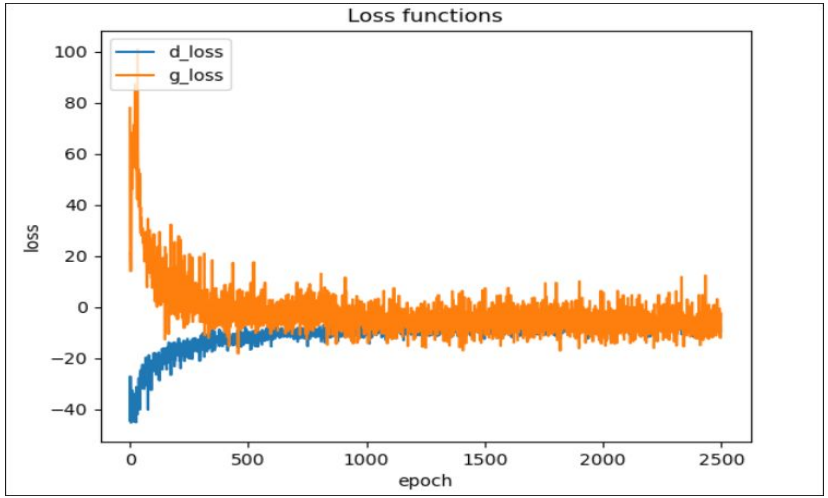
Vantaggi della WCGAN-GP:

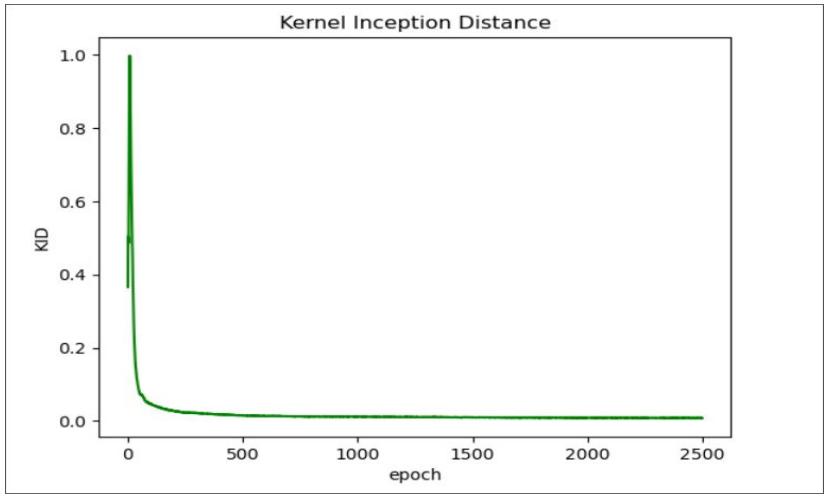
- Distanza di Wasserstein
- Gradient Penalty
- Modulo condizionale

Fase di addestramento e valutazione

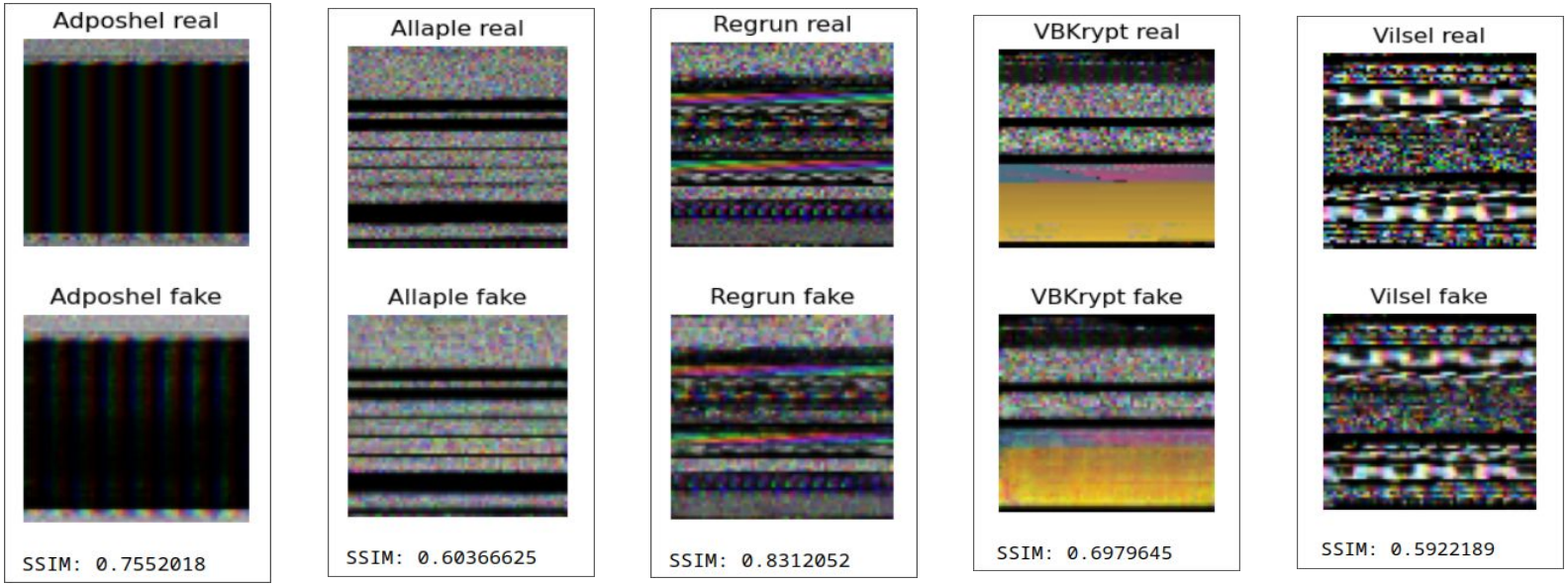
Risultati finali:

- Perdita del discriminatore (d_loss) = −6.7783
- Perdita del generatore (g_loss) = −6.6646
- KID (Kernel Inception Distance) = 0.008





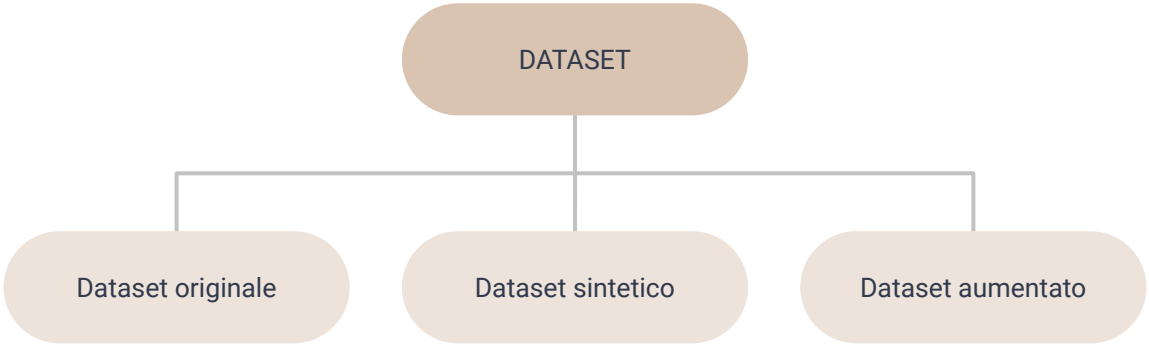
Fase di generazione



SSIM (Structural Similarity Index Metric)

Classificazione multiclasse

È stata utilizzata una ResNet152 (Residual Network con 152 livelli) come classificatore di malware per valutare l'efficacia della Data Augmentation con dati sintetici.



Esperimenti con ResNet152

Esperimento 1	Esperimento 2	Esperimento 3	Esperimento 4
Addestramento e valutazione con dataset originale	Addestramento con dataset originale e valutazione con dataset sintetico	Addestramento con dataset aumentato e valutazione con dataset sintetico	Addestramento e valutazione con dataset aumentato

Metriche di valutazione e risultati

1° esperimento

Loss	Accuracy	Precision	Recall	F1-Score
0.0009	1	1	1	1

2° esperimento

Loss	Accuracy	Precision	Recall	F1-Score
1.04	0.76	0.80	0.74	0.77

3° esperimento

Loss	Accuracy	Precision	Recall	F1-Score
0.08	0.97	0.98	0.97	0.97

4° esperimento

Loss	Accuracy	Precision	Recall	F1-Score
0.01	0.99	0.99	0.99	0.99

Conclusioni

1. Generazione di malware sintetici
2. Data augmentation con dati sintetici
3. Ri-addestramento del classificatore di malware
4. Migliori prestazioni nel rilevamento di malware zero-day

Sviluppi futuri

- WCGAN-GP più complessa e immagini di malware di dimensioni maggiori
- Varianti di GAN o altri modelli generativi
- Utilizzo di altri tipi di dataset di malware
- Conversione delle immagini generate in codice

ISLab
Intelligent
Systems
Laboratory

Grazie per l’attenzione