Davide Maiorca - Intervista al Forum ICT Security 2018

Author : Redazione

Date : 15 novembre 2018

Davide Maiorca

Ricercatore Post-Doc in Sicurezza Informatica presso l'Università degli Studi di Cagliari

Diversi studi nel campo del Machine Learning e della Sicurezza Informatica hanno dimostrato che anche gli algoritmi più avanzati di Intelligenza Artificiale possono essere ingannati. Uno degli ultimi esempi in tal senso ha visto un attaccante modificare un'immagine attraverso delle manipolazioni mirate, portando l'Intelligenza Artificiale a classificarla in modo molto diverso rispetto a quello che era l'immagine iniziale. Questo ovviamente può essere utilizzato anche per violare la sicurezza informatica tramite dei malware modificati.

In merito al malware, l'attaccante può prendere un file che viene riconosciuto come maligno e modificare alcune sue proprietà, ovviamente in modo mirato seguendo determinati algoritmi, e con pochi accorgimenti può violare un sistema di difesa anche in modo devastante.

All'Università di Cagliari studiamo come rendere più sicuri i sistemi di rilevazione; sostanzialmente questo si può fare in due modi: il primo cercando informazioni che siano difficili da modificare, e ovviamente non è semplice all'atto pratico; il secondo intervenendo tramite algoritmi di rilevazione quindi prendendone di già esistenti e ben funzionanti e modificandoli in modo tale da incrementare la resistenza agli attacchi. Ciò comporta una specie di trade-off fra la precisione del sistema e la sua robustezza. L'obiettivo finale è quello di garantire una buona affidabilità sia contro gli attacchi standard sia che contro quelli evasivi.

Domande:

1. L'Intelligenza Artificiale può essere "ingannata"?

2. Come si può usare il malware per aggirare i sistemi di rilevazione basati sull’Intelligenza Artificiale?
3. È possibile rendere più sicuri i sistemi di rilevazione?

Vai all'evento Forum ICT Security 2018