ε.δε.μ²
ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ & ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

infostrag

Η ισχύς εν τη ενώσει: Συνεργατική Ανίχνευση & Αντιμετώπιση Κυβερνο-επιθέσεων Distributed Denial of Service (DDoS) με Ομόσπονδες Αρχιτεκτονικές Τεχνητής Νοημοσύνης - Federated Learning

Βασίλης Μάγκλαρης

Ομότιμος Καθηγητής Σχολής Ηλεκτρολόγων Μηχ. & Μηχ. Υπολογιστών Ε.Μ.Π.

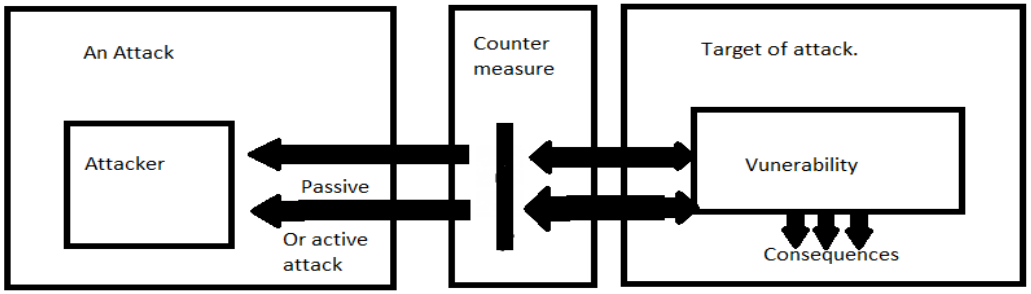
maglaris@netmode.ntua.gr www.netmode.ntua.gr

18ο Σεμινάριο της Ερμούπολης για την Κοινωνία της Πληροφορίας & την Οικονομία της Γνώσης

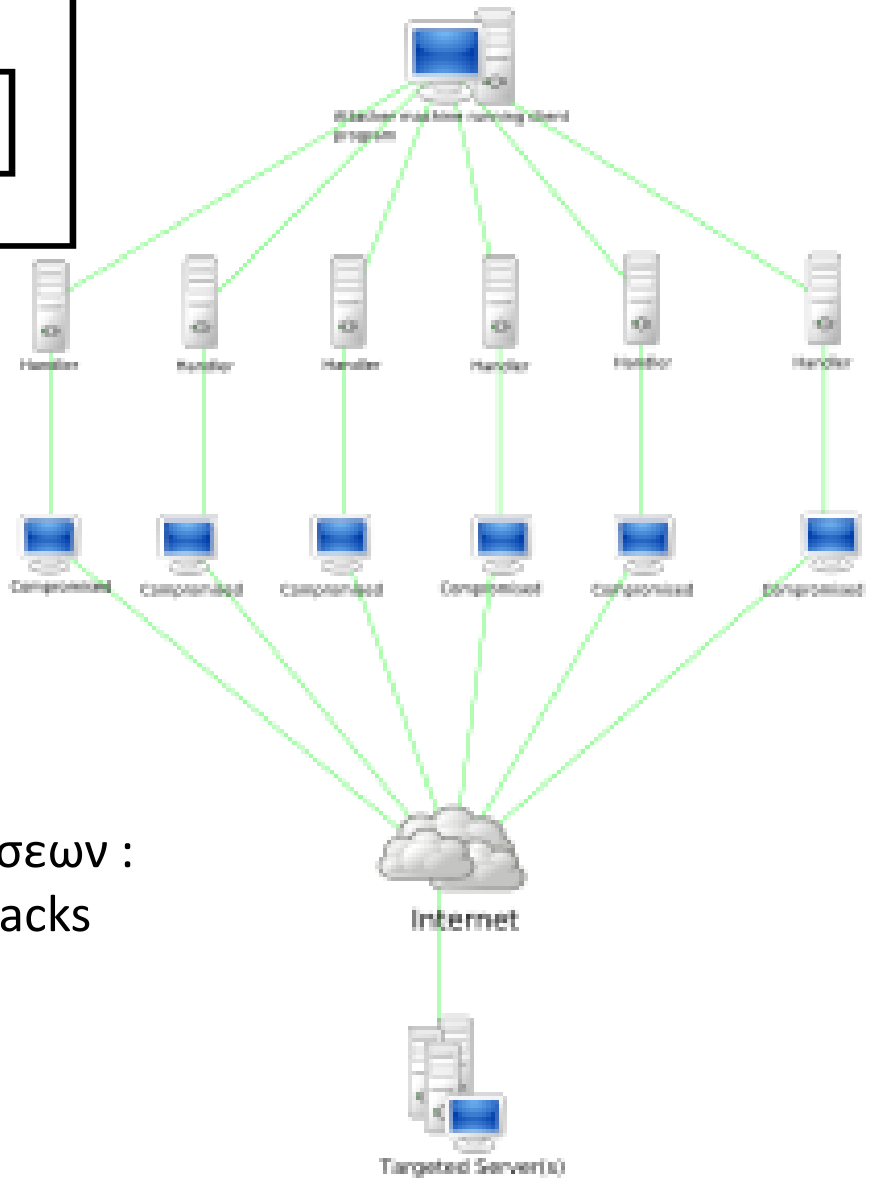
Συνεδριακή Αίθουσα Επιμελητηρίου Κυκλάδων
Παρασκευή 14/7/2023

Κυβερνο-επιθέσεις &
Distributed Denial of Service (DDoS) Attacks

Use-case: Ανίχνευση Κυβερνο-επιθέσεων (Cyberattacks) στο Internet



https://en.wikipedia.org/wiki/Cyberattack



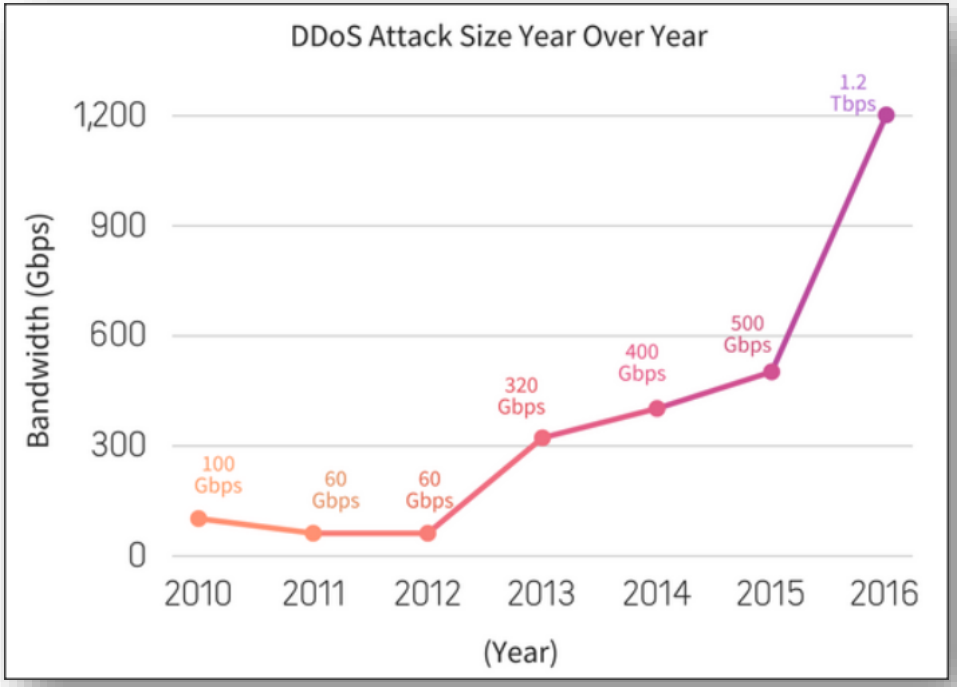
Πιο συχνό είδος ενεργών (active) επιθέσεων :
Distributed Denial of Service (**DDoS**) attacks

Use-case: Ανίχνευση Κυβερνο-επιθέσεων (Cyberattacks) στο Internet

Επίθεση DDoS σε Πάροχο Υπηρεσιών DNS Dyn (21/10/2016)

- Μέγεθος κίνησης: **1.2 Tbps**
- Πηγή της επίθεσης **100.000** παραβιασμένες συσκευές Internet of Things
- Αδυναμία πρόσβασης μεγάλου αριθμού χρηστών σε σημαντικές υπηρεσίες επιχειρήσεων: **Amazon, CNN, Twitter, PayPal, Visa, GitHub, Spotify, Netflix,…**

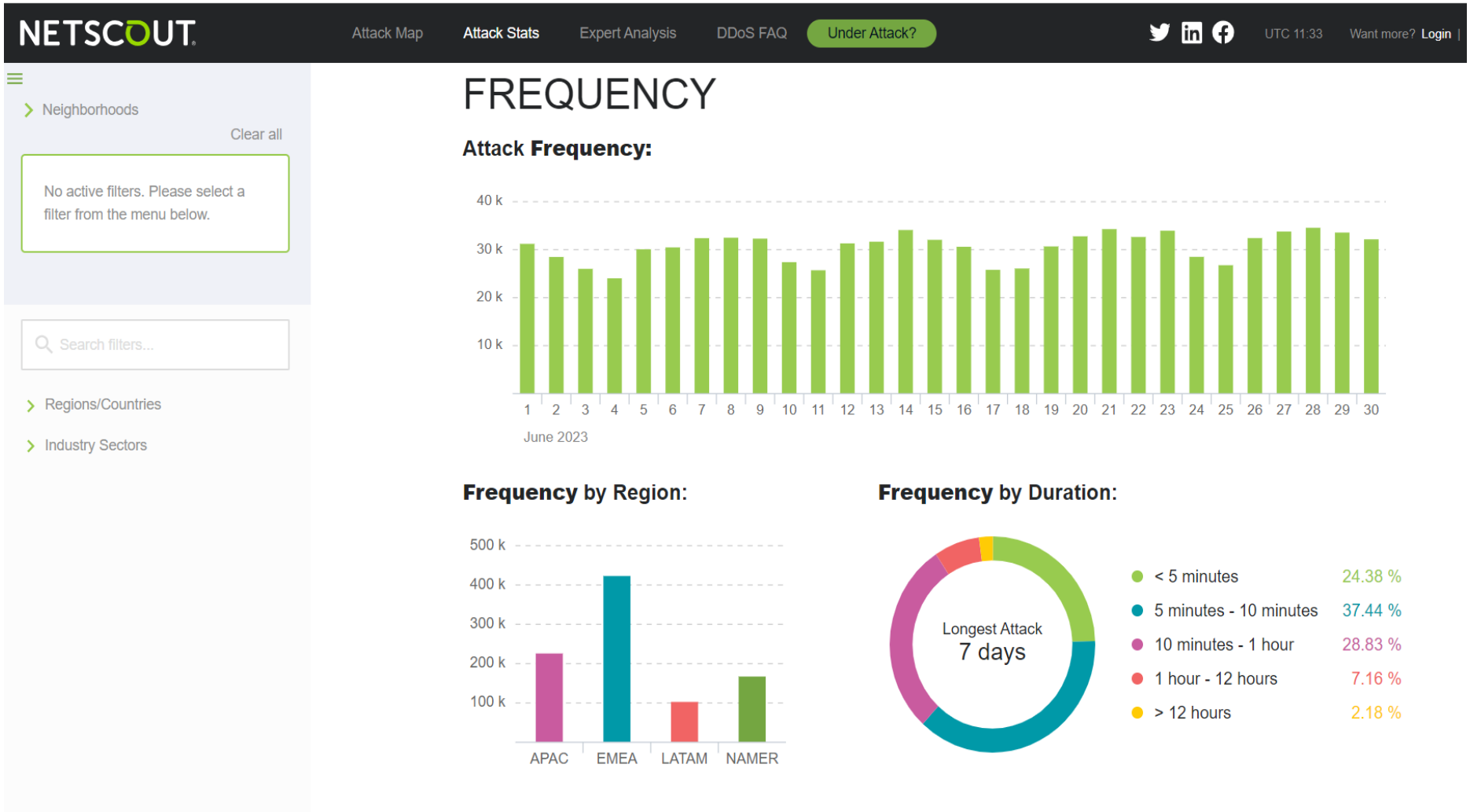
https://blogs.haltdos.com/wp-content/uploads/2017/02/2015.png



Use-case: Ανίχνευση Κυβερνο-επιθέσεων (Cyberattacks) στο Internet

June 2023 DDoS Summary (1/3)

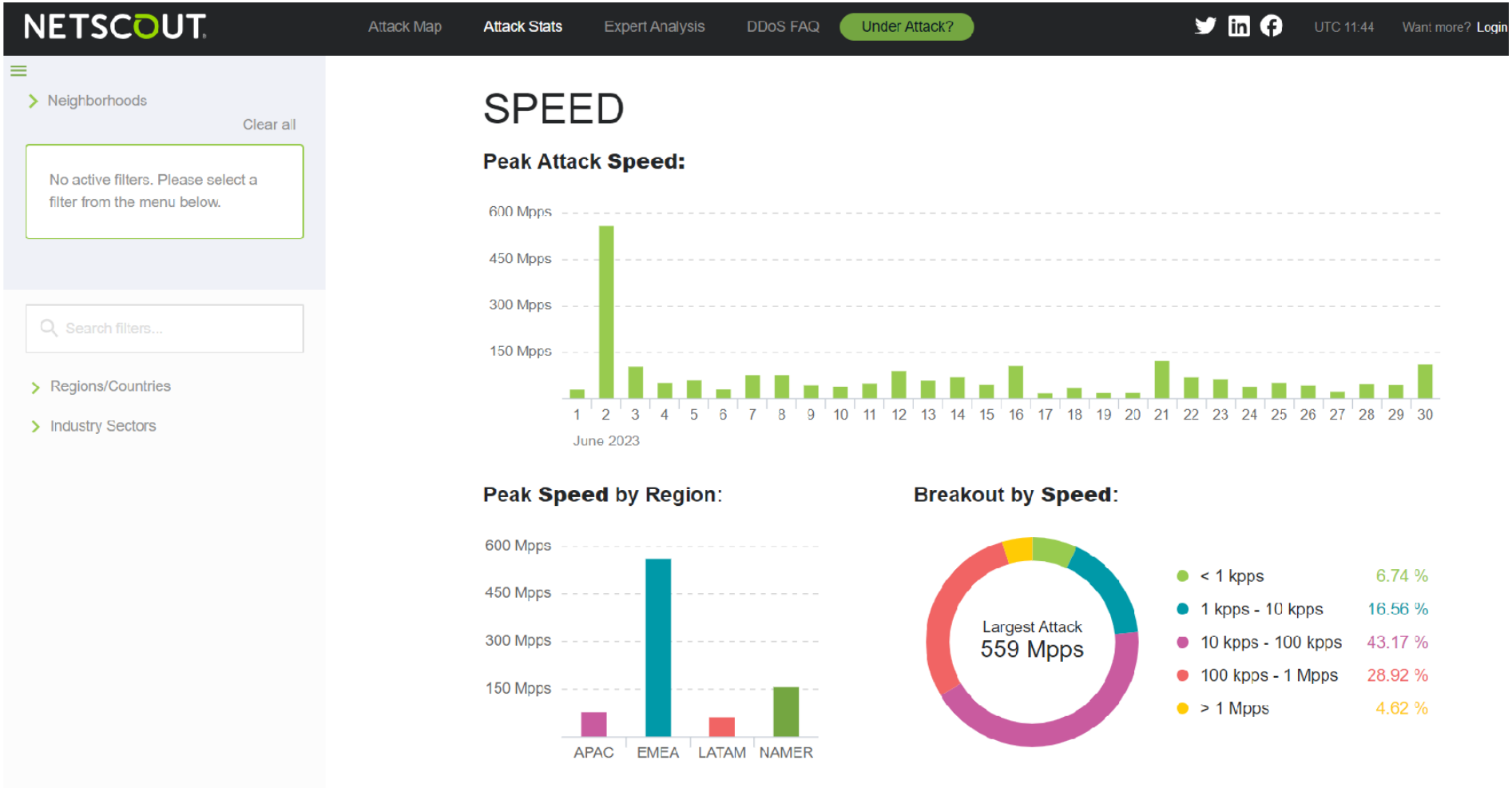
https://horizon.netscout.com/?atlas=summary



Use-case: Ανίχνευση Κυβερνο-επιθέσεων (Cyberattacks) στο Internet

June 2023 DDoS Summary (2/3)

https://horizon.netscout.com/?atlas=summary



Use-case: Ανίχνευση Κυβερνο-επιθέσεων (Cyberattacks) στο Internet

June 2023 DDoS Summary (3/3)

https://horizon.netscout.com/?atlas=summary

NETSCOUT

Attack Map | Attack Stats | Expert Analysis | DDoS FAQ | Under Attack?

Neighborhoods

Clear all

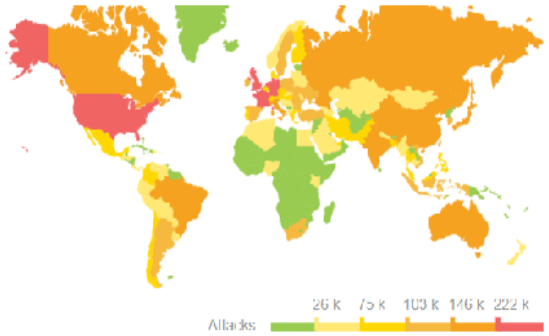
No active filters. Please select a filter from the menu below.

Search filters...

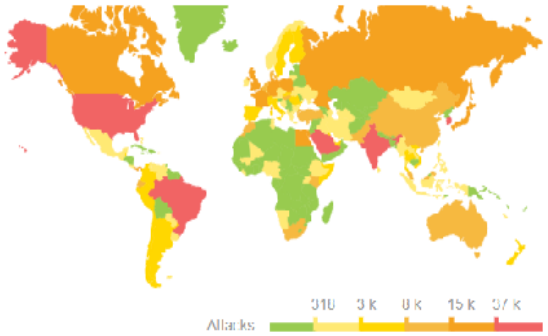
Regions/Countries

Industry Sectors

Top **Source** Countries:



Top **Destination** Countries:



Sources

United States	486,717	53 %
Germany	234,578	25.5 %
United Kingdom	232,403	25.3 %
Netherlands	225,153	24.5 %
France	221,815	24.2 %

Destinations

United States	184,900	20.1 %
Brazil	59,424	6.5 %
South Korea	45,074	4.9 %
India	39,523	4.3 %
Saudi Arabia	36,959	4 %

Ανίχνευση (Identification) & Αντιμετώπιση (Mitigation) Κυβερνο-επιθέσεων

Μηχανισμοί Ανίχνευσης Επιθέσεων με Ευφυείς Αλγορίθμους

- Ανίχνευση μέσω *monitoring* → εντοπισμός ξαφνικών αλλαγών κίνησης (*outlier detection*)
- Χρήση εργαλείων Τεχνητής Νοημοσύνης (*Artificial Intelligence* - *AI*)
- Έξυπνοι αλγόριθμοι Μηχανικής Μάθησης (*Machine Learning* - *ML*) με διαμόρφωση και παραμέτρους ρυθμιζόμενες βάση της εμπειρίας πρόσφατων επιθέσεων (εκτενή αρχεία μάθησης – *learning datasets*)
- Ανάλυση ακρίβειας προβλέψεων συστημάτων *ML*, δυνατότητα γενίκευσης (*generalization*) πέραν του δείγματος μάθησης (*learning sample*) για νέες επιθέσεις εκτίμηση με χρήση διαφορετικών δειγματικών στοιχείων (*test sample elements*)
- Διαθεσιμότητα αξιόπιστων αρχείων από επιθέσεις, ζητήματα ιδιωτικότητας (*privacy*) προσωπικών/εταιρικών δεδομένων → φρένο στην συνεργατική έρευνα – καινοτομία (;;;)
- Ερμηνεία επιλογών (*eXplainable AI* – *XAI*) για χρήστες/διαχειριστές & ελεγκτικές/ρυθμιστικές αρχές (νέο πεδίο έρευνας)

Ανίχνευση (Identification) & Αντιμετώπιση (Mitigation) Κυβερνο-επιθέσεων

Μηχανισμοί Ανίχνευσης Επιθέσεων με Ευφυείς Αλγορίθμους

- Ανίχνευση μέσω *monitoring* → εντοπισμός ξαφνικών αλλαγών κίνησης (*outlier detection*)
- Χρήση εργαλείων Τεχνητής Νοημοσύνης (*Artificial Intelligence* - *AI*)
- Έξυπνοι αλγόριθμοι Μηχανικής Μάθησης (*Machine Learning* - *ML*) με διαμόρφωση και παραμέτρους ρυθμιζόμενες βάση της εμπειρίας πρόσφατων επιθέσεων (εκτενή αρχεία μάθησης – *learning datasets*)
- Ανάλυση ακρίβειας προβλέψεων συστημάτων *ML*, δυνατότητα γενίκευσης (*generalization*) πέραν του δείγματος μάθησης (*learning sample*) για νέες επιθέσεις εκτίμηση με χρήση διαφορετικών δειγματικών στοιχείων (*test sample elements*)
- Διαθεσιμότητα αξιόπιστων αρχείων από επιθέσεις, ζητήματα ιδιωτικότητας (*privacy*) προσωπικών/εταιρικών δεδομένων → φρένο στην συνεργατική έρευνα – καινοτομία (;;;)
- Ερμηνεία επιλογών (*eXplainable AI* – *XAI*) για χρήστες/διαχειριστές & ελεγκτικές/ρυθμιστικές αρχές (νέο πεδίο έρευνας)

Μηχανισμοί Αντιμετώπισης Επιθέσεων

- Διαμόρφωση φίλτρων (*firewalls*) μπλοκαρίσματος κακόβουλης κίνησης (βάση ύποπτων ροών πακέτων – *flows* ή ύποπτων υπογραφών σε πακέτα – *packet signatures* ή υπόπτων διευθύνσεων πηγής…)
- Ζητήματα ακρίβειας, ταχύτητας παρέμβασης, κλιμάκωσης, αντιμετώπισης *spoofed addresses*, συνεργατικών παρεμβάσεων με *upstream transient networks*…
- Χρήση γρήγορων φίλτρων στο *data plane* (προγραμματιζόμενες *XDP* κάρτες δικτύου, *deep programmable* μεταγωγείς σε γλώσσα *P4*…)

Τεχνητή Νοημοσύνη & Μηχανική Μάθηση

Εισαγωγικά περί Μηχανικής Μάθησης (1/5)

Διασχολικό Μεταπτυχιακό Πρόγραμμα Σπουδών Ε.Μ.Π.

Επιστήμη Δεδομένων – Μηχανική Μάθηση, Data Science – Machine Learning

https://dsml.ece.ntua.gr/

Ορισμοί

Τεχνητή Νοημοσύνη (Artificial Intelligence - AI):
Artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind (IBM: https://www.ibm.com/topics/artificial-intelligence)

Μηχανική Μάθηση (Machine Learning - ML):
Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy (IBM: https://www.ibm.com/topics/machine-learning)

Εισαγωγικά περί Μηχανικής Μάθησης (2/5)

Λόγοι ανάπτυξης Μηχανικής Μάθησης

- Η κατακλυσμιαία ανάπτυξη υπολογιστικών υποδομών αποθήκευσης και επεξεργασίας δεδομένων, επιτρέπει σήμερα την υλοποίηση αλγορίθμων στατιστικής ανάλυσης και στοχαστικής βελτιστοποίησης με βάση ιστορικά στοιχεία δείγματος μάθησης

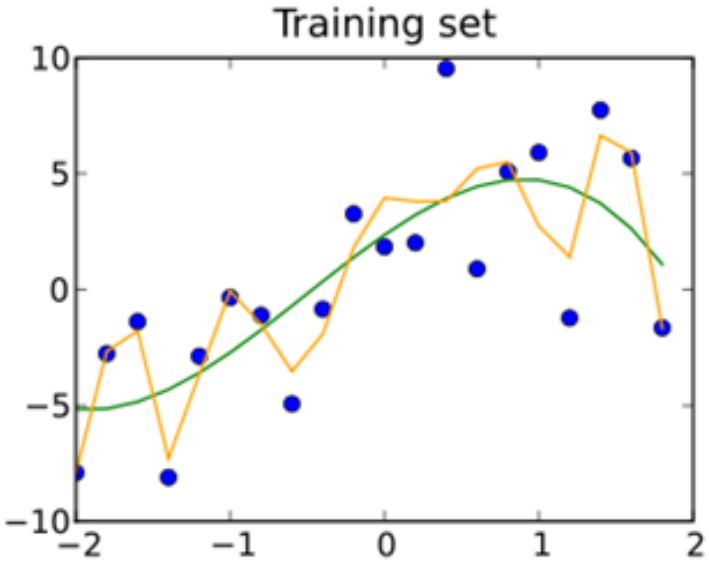
- Η αλματώδης συσσώρευση τεράστιου όγκου πολυδιάστατων δεδομένων (*big data*) με πολλά χαρακτηριστικά, απαιτεί την ανάπτυξη ευφυών αλγορίθμων εξόρυξης εκτιμήσεων, προβλέψεων και ταξινόμησης νεοεμφανιζόμενων δειγματικών στοιχείων

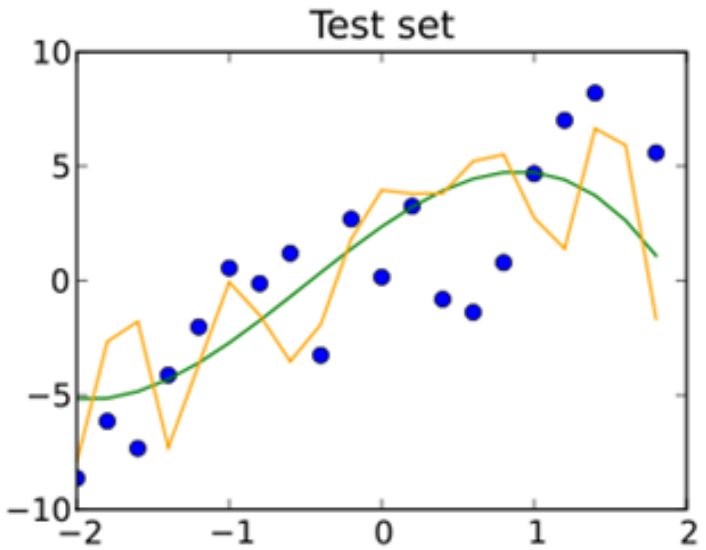
- Η κατανόηση μεθόδων μάθησης σε βιολογικά συστήματα οδηγεί σε αλγορίθμους τεχνητής νοημοσύνης για συμπλήρωση και ελεγχόμενη πρόβλεψη (ή/και δημιουργία) δειγματικών στοιχείων, συμπεριλαμβανομένων ακολουθιών και χρονοσειρών (στοχαστικών διαδικασιών, *stochastic processes*) με βάση παρεμφερή στατιστικά χαρακτηριστικά αποθηκευμένου δείγματος μάθησης

Εισαγωγικά περί Μηχανικής Μάθησης (3/5)

Ορισμοί Συνόλων Δεδομένων (Datasets)

https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets





Training Dataset (***μπλε σημεία μάθησης***)

- Λεπτομερής **κίτρινη** καμπύλη εκτίμησης με απόκλιση MSE=4
- Απλή **πράσινη** καμπύλη με απόκλιση MSE=9

Test Dataset (***μπλε σημεία γενίκευσης***)

- Απόκλιση από **κίτρινη** καμπύλη MSE=15 (από 4) ***OVERFITTING***
- Απόκλιση από **πράσινη** καμπύλη MSE=13 (από 9)

Εισαγωγικά περί Μηχανικής Μάθησης (4/5)

Μοντέλα Μηχανικής Μάθησης

Διακριτικά Μοντέλα (**Discriminative Models**):

Μέθοδοι ταξινόμησης (*classification*) ή εκτίμησης (παλινδρόμηση, *regression*) δειγματικών στοιχείων (*data elements*) μέσω υπό συνθήκη πιθανότητας (*conditional density*) εξόδου (*label*) βάσει χαρακτηριστικών (*features*) του, όπως αυτές προσεγγίστηκαν σε στοιχεία δείγματος μάθησης (*training sample*) για γενίκευση σε *test datasets* (*generalization*)

Ενδεικτικές Εφαρμογές:

- *Ταξινόμηση δειγματικών στοιχείων* με βάση συνάρτηση χαρακτηριστικών τους
- *Αναγνώριση προτύπων* με βάση κύρια χαρακτηριστικά τους (*pattern recognition*)
- *Εκτίμηση εξόδου* συμβατή με διαθέσιμα ζεύγη εισόδου - στόχου (*regression*)

Παραγωγικά Μοντέλα (**Generative Models**):

Μέθοδοι εκτίμησης τρόπων παραγωγής (*generation*) δειγματικών στοιχείων, στατιστικά συμβατών με ιδιότητες του δείγματος μάθησης (*training sample*) μέσω συνδυασμένων πιθανοτήτων (*joint probabilities*) εξόδου (*output*) και χαρακτηριστικών (*features*) εισόδου, όπως υπολογίστηκαν στα στοιχεία του δείγματος μάθησης (*training ample elements*)

Ενδεικτικές Εφαρμογές:

- *Δημιουργία προσομοιωμένων στοιχείων*: κειμένων (συμβατών με αποδεκτά μοντέλα Natural Language processing - NLP), εικόνων, κινούμενων σχεδίων, ιδεατών τοπίων…
- *Εμπλουτισμός Μηχανών Αναζήτησης* (*Google*, *MS Bing* + *OpenAI Chat Generative Pre-trained Transformer - ChatGPT*)
- *Επικράτηση αληθοφανών εναλλακτικών εκτιμήσεων* σε συνέργεια με εργαλεία θεωρίας παιγνίων (*Generative Adversarial Networks – GAN*)

Εισαγωγικά περί Μηχανικής Μάθησης (5/5)

Γενικές Κατηγορίες Συστημάτων Μηχανικής Μάθησης

- **Επιβλεπόμενη Μάθηση με Εκπαιδευτή - Supervised Learning**
  - Χρήση δεδομένων μάθησης με συνημμένες επιθυμητές αποκρίσεις εξόδου (*labeled training sample points*) που εκπαιδεύουν σε πρώτη φάση το σύστημα Μηχανικής Μάθησης μέσω εξωτερικού εκπαιδευτή για αναζήτηση απόκρισης (ταξινόμηση, πρόβλεψη) σε επόμενη φάση γενίκευσης με νέα δεδομένα εισόδου

Εισαγωγικά περί Μηχανικής Μάθησης (5/5)

Γενικές Κατηγορίες Συστημάτων Μηχανικής Μάθησης

- **Επιβλεπόμενη Μάθηση με Εκπαιδευτή - Supervised Learning**
  - Χρήση δεδομένων μάθησης με συνημμένες επιθυμητές αποκρίσεις εξόδου (*labeled training sample points*) που εκπαιδεύουν σε πρώτη φάση το σύστημα Μηχανικής Μάθησης μέσω εξωτερικού εκπαιδευτή για αναζήτηση απόκρισης (ταξινόμηση, πρόβλεψη) σε επόμενη φάση γενίκευσης με νέα δεδομένα εισόδου
- **Μάθηση χωρίς Εκπαιδευτή**
  - Μη Επιβλεπόμενη Μάθηση - **Unsupervised Learning** όπου το σύστημα αυτορυθμίζεται ανακαλύπτοντας από μόνο του ενδιαφέρουσες στατιστικές δομές (*stochastic features, patterns*) σε μεγάλο όγκο μη χαρακτηρισμένων δεδομένων (*unlabeled datasets*) ώστε να προκύπτουν μοντέλα, μέθοδοι επεξεργασίας, αποθήκευσης και ταξινόμησής, π.χ. σε ομάδες (*clusters*)
  - Ενισχυτική Μάθηση - **Reinforcement Learning** όπου το σύστημα αντιδρά σε σήματα επιβράβευσης/αποθάρρυνσης μέσω *agents* από το περιβάλλον εισόδου, προς το οποίο κοινοποιεί ενέργειές του (*actions*) που επηρεάζουν την εξέλιξη της κατάστασης του περιβάλλοντος για την επίτευξη μακροπρόθεσμου στόχου

Εισαγωγικά περί Μηχανικής Μάθησης (5/5)

Γενικές Κατηγορίες Συστημάτων Μηχανικής Μάθησης

- **Επιβλεπόμενη Μάθηση με Εκπαιδευτή - Supervised Learning**
  - Χρήση δεδομένων μάθησης με συνημμένες επιθυμητές αποκρίσεις εξόδου (*labeled training sample points*) που εκπαιδεύουν σε πρώτη φάση το σύστημα Μηχανικής Μάθησης μέσω εξωτερικού εκπαιδευτή για αναζήτηση απόκρισης (ταξινόμηση, πρόβλεψη) σε επόμενη φάση γενίκευσης με νέα δεδομένα εισόδου
- **Μάθηση χωρίς Εκπαιδευτή**
  - Μη Επιβλεπόμενη Μάθηση - **Unsupervised Learning** όπου το σύστημα αυτορυθμίζεται ανακαλύπτοντας από μόνο του ενδιαφέρουσες στατιστικές δομές (*stochastic features, patterns*) σε μεγάλο όγκο μη χαρακτηρισμένων δεδομένων (*unlabeled datasets*) ώστε να προκύπτουν μοντέλα, μέθοδοι επεξεργασίας, αποθήκευσης και ταξινόμησής, π.χ. σε ομάδες (*clusters*)
  - Ενισχυτική Μάθηση - **Reinforcement Learning** όπου το σύστημα αντιδρά σε σήματα επιβράβευσης/αποθάρρυνσης μέσω *agents* από το περιβάλλον εισόδου, προς το οποίο κοινοποιεί ενέργειές του (*actions*) που επηρεάζουν την εξέλιξη της κατάστασης του περιβάλλοντος για την επίτευξη μακροπρόθεσμου στόχου

Η επιβλεπόμενη μάθηση προσφέρει απόδοση, αξιοπιστία και ταχύτητα για προβλήματα που αφορούν σε αποφάσεις χειρισμού δεδομένων μετά από διαδικασία μάθησης αλλά απαιτεί **labeled learning data sets** που δεν είναι εύκολα διαθέσιμα

Εισαγωγικά περί Μηχανικής Μάθησης (5/5)

Γενικές Κατηγορίες Συστημάτων Μηχανικής Μάθησης

- **Επιβλεπόμενη Μάθηση με Εκπαιδευτή - Supervised Learning**
  - Χρήση δεδομένων μάθησης με συνημμένες επιθυμητές αποκρίσεις εξόδου (*labeled training sample points*) που εκπαιδεύουν σε πρώτη φάση το σύστημα Μηχανικής Μάθησης μέσω εξωτερικού εκπαιδευτή για αναζήτηση απόκρισης (ταξινόμηση, πρόβλεψη) σε επόμενη φάση γενίκευσης με νέα δεδομένα εισόδου
- **Μάθηση χωρίς Εκπαιδευτή**
  - Μη Επιβλεπόμενη Μάθηση - **Unsupervised Learning** όπου το σύστημα αυτορυθμίζεται ανακαλύπτοντας από μόνο του ενδιαφέρουσες στατιστικές δομές (*stochastic features, patterns*) σε μεγάλο όγκο μη χαρακτηρισμένων δεδομένων (*unlabeled datasets*) ώστε να προκύπτουν μοντέλα, μέθοδοι επεξεργασίας, αποθήκευσης και ταξινόμησής, π.χ. σε ομάδες (*clusters*)
  - Ενισχυτική Μάθηση - **Reinforcement Learning** όπου το σύστημα αντιδρά σε σήματα επιβράβευσης/αποθάρρυνσης μέσω *agents* από το περιβάλλον εισόδου, προς το οποίο κοινοποιεί ενέργειές του (*actions*) που επηρεάζουν την εξέλιξη της κατάστασης του περιβάλλοντος για την επίτευξη μακροπρόθεσμου στόχου

Η επιβλεπόμενη μάθηση προσφέρει απόδοση, αξιοπιστία και ταχύτητα για προβλήματα που αφορούν σε αποφάσεις χειρισμού δεδομένων μετά από διαδικασία μάθησης αλλά απαιτεί **labeled learning data sets** που δεν είναι εύκολα διαθέσιμα

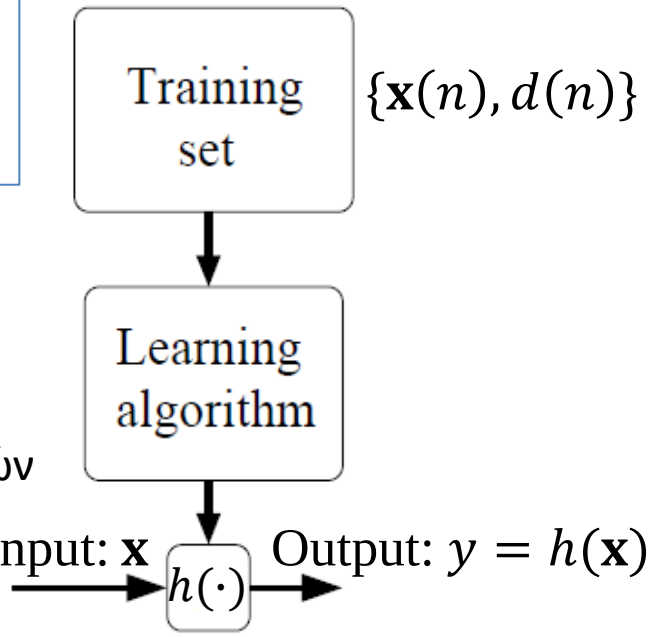
- **Ομόσπονδη Μάθηση - Federated Learning**
  - Συνεργατική μάθηση με κοινό μοντέλο → «μέσο όρο» μοντέλων που προτείνουν οι συμμετέχοντες σε *multi-domain* ομοσπονδία **βάσει τοπικών δεδομένων μάθησης**

Γενικό Μοντέλο Επιβλεπόμενης Μάθησης - Supervised Learning

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018

- Στόχος του συστήματος είναι η αντιστοίχηση ενός δειγματικού στοιχείου εισόδου (***input sample point***, ***example***, ***instance***) $\mathbf{x} = [x_1\ x_2\ \dots x_m]^{\mathrm{T}}$ σε τιμές εξόδου y που εκτιμούν επιθυμητές τιμές d (***labels***, ***targets***) π.χ. πρόβλεψη ή ταξινόμηση. Τα στοιχεία x_i είναι αριθμητικές τιμές που κωδικοποιούν m ειδοποιά χαρακτηριστικά (***features***) του δειγματικού στοιχείου $\mathbf{x}$

Ζητείται ο προσδιορισμός της συνάρτησης εισόδου - εξόδου $y = h(\mathbf{x}) \cong d$ που προκύπτει από δείγμα μάθησης (***Training Set***) N ***labeled*** ζευγών $\{\mathbf{x}(n), d(n)\}$, $n = 1,2, \dots, N$ γνωστών σε εξωτερικό εκπαιδευτή (***supervisor***)



- Η μορφή και οι παράμετροι της $h(\cdot)$ προσδιορίζονται με αλγόριθμο μάθησης που συγκλίνει σε προσέγγιση του στόχου της υπόθεσης για τα N στοιχεία του δείγματος μάθησης

$$d(n) \cong y(n) = h(\mathbf{x}(n))$$

- Αν ο στόχος ικανοποιείται με μικρό αριθμό διακριτών επιλογών (κλάσεων) της y πρόκειται για πρόβλημα Ταξινόμησης, ***Classification*** (για δύο κλάσεις έχουμε δυαδική ταξινόμηση)
- Αν η έξοδος y λαμβάνει συνεχείς τιμές, το πρόβλημα αναφέρεται σαν Παλινδρόμηση, ***Regression***

DDoS Attack Detection via Privacy-aware Federated Learning and Collaborative Mitigation in Multi-domain Cyber Infrastructures

Marinos Dimolianis, Dimitrios Kalogeras, Nikos Kostopoulos & Vasilis Maglaris

2022 IEEE 11th Conference on Cloud Networking (CloudNet)
(best Paper Award)

Limitations in Collaborative Detection/Mitigation

- **Collaborative Detection**

  - Hindered by data privacy legislations, e.g. **GDPR**
  - Slow-paced human-driven procedures for inter-domain collaboration

- **Collaborative Mitigation**

  - Limited filtering capabilities of legacy firewall solutions
    → Blackholing
    → IP-based Filtering Rules
    → Flow-based rules (BGP FlowSpec)

Legacy firewalls are typically less effective than signature-based rules that combine data packet fields of multiple protocol layers

Introduction

Collaborative DDoS protection for interconnected cyber infrastructures

Key Contributions:

- Collaborative DDoS Detection via Federated Learning

  Collaborators converge on federated Machine Learning (ML) models without sharing any private domain-specific data

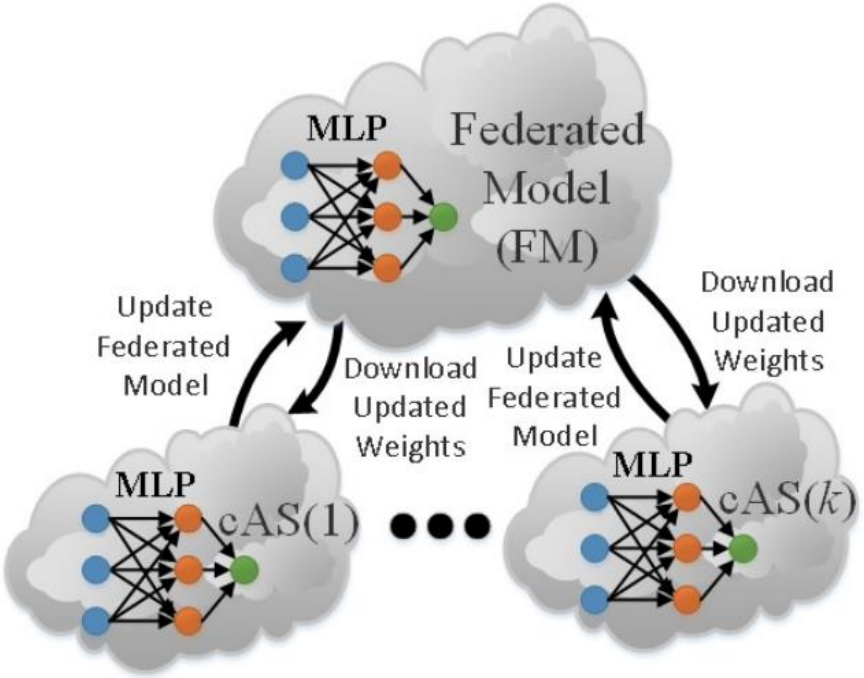
- DDoS Mitigation via cloud-native and scalable programmable firewalls based on the XDP data plane framework

  Matching and blocking arbitrary packet field combinations (signatures) based on programmable data plane firewalls

- Cross-domain propagation of DDoS filtering requests

  BGP URI signaling to disseminate application filtering rules across multi-domain environments

Background: Federated Learning



Iterative Process:

- Collaborating parties train ML models, e.g. supervised learning Multi-layer Perceptrons (MLP's)
- Model weights are exported to a Third-Trusted Party (TTP)
- TTP aggregates collaborator models into a Federated Model (FM)
- FM is conveyed to participants for validation

Cross-silo (multi-domain) FL:

Participants are:

- Moderate in number, e.g. data centers
- Highly available
- Fault tolerant

Cross-device FL: *Introduced by Google* (2016)

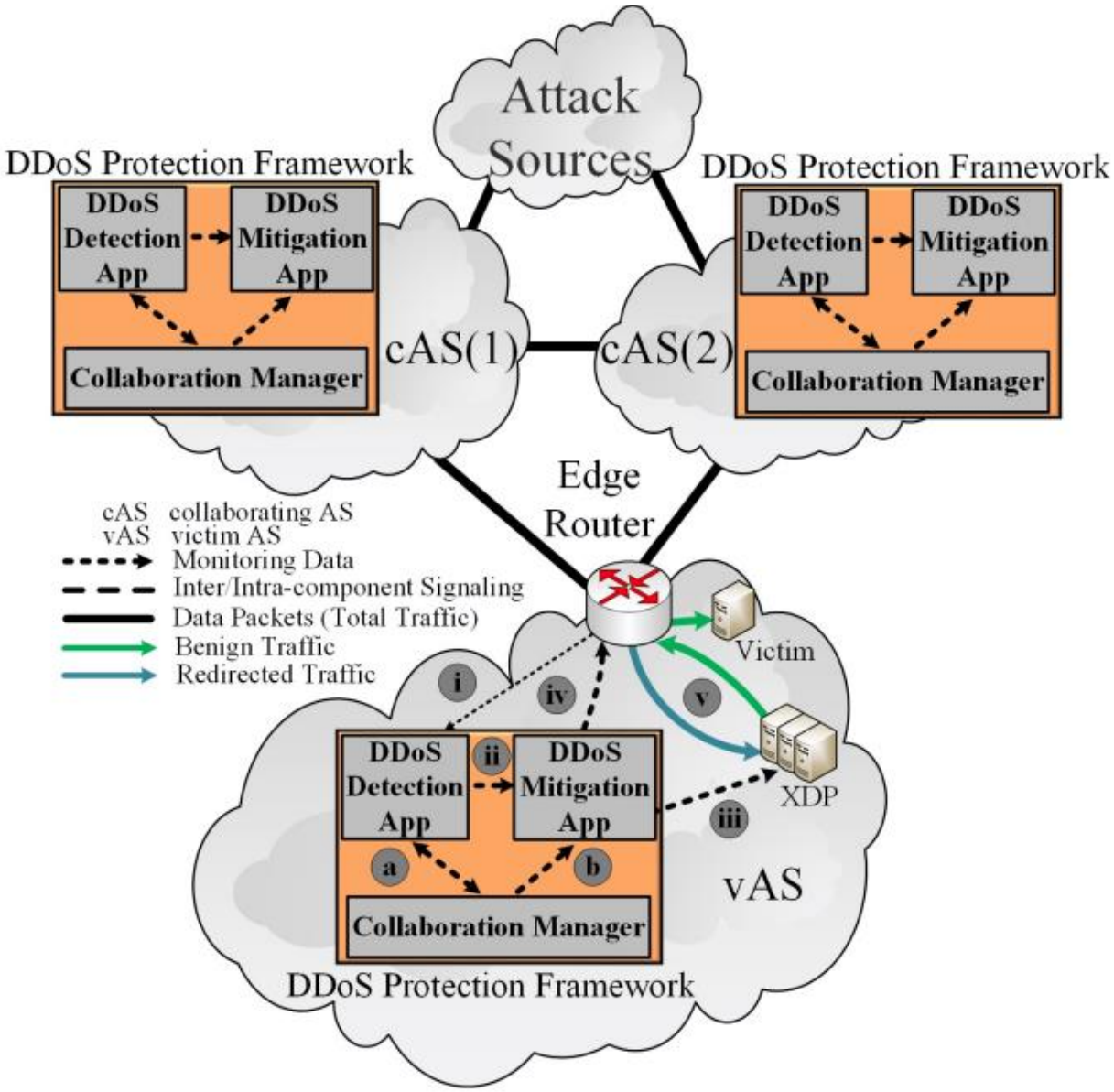
Participants are:

- Large in number, e.g. smartphones
- Possibly unreliable/untrusted

High-Level Description

Distributed Privacy-Aware ML via Cross-silo Federated Learning (FL)

- Trusted collaboration of knowledgeable, independently managed participants (Autonomous Systems – AS's, Datacenters)
- Participants do not share sensitive datasets for training
- Participants adopt a common ML model



DDoS Detection App (1/2)

- Retrieves packet-based data from external monitoring mechanisms
- Aggregates data based on preselected packet fields and specific time-windows to form packet signatures per victim (sub)network
- Identifies malicious signatures using machine learning models

Federated Learning for malicious signature detection:

- **Privacy-awareness:** No benign/attack data are exchanged
- **Bandwidth conservation:** Model weights exchanged instead of data

Prerequisites for collaboration:

- Agreement on common machine learning model architectures and training methods (e.g. features, number of neurons, …)
- Coordination by a third-trusted party, e.g. a major Internet Exchange (IX)

DDoS Detection App (2/2)

- **Selection of common local model:** Multi-Layer Perceptrons (MLP's) for binary classification
- MLP's are trained in multiple **local epochs** and **FL rounds**
  - **Local epochs:** One training iteration over a collaborator's data
  - **FL rounds:** total aggregations of local weights
- **Federated Averaging for weight aggregation:** Weighted mean of collaborator weights
- **Training termination:** Once the (weighted) average classification accuracy reaches an acceptable level

$$w_{FM} = \sum_{i=1}^{k} \frac{N_i}{N} w_i, \; where \; N = \sum_{i=1}^{k} N_i$$

Training examples of participant i

FM weights

Total training examples

Weights of local model i

Background: The eXpress Data Path (XDP) Framework

- Establishes a programmable data path in the Linux Kernel

- **XDP Hook:**

- Ingress traffic detained & processed before any memory allocation
- Delivers packets to an extended Berkeley Packet Filter (**eBPF**) Program

- **eBPF Verifier:** Ensures that XDP will not compromise kernel safety:
  - No unbounded loops present
  - Maximum eBPF program size limited
  - No data are read out of bounds

- **eBPF Maps:** Data structures used for communication among user space and eBPF programs

Note: No requirements for specific hardware, e.g. **P4** switches

DDoS Mitigation App (1/2)

Receives filtering requests for ongoing attacks and raises appropriate mitigation countermeasures

- **Filtering rules:** Arbitrary packet field combinations (**signatures**)
- Signatures are stored in **eBPF** maps (flexibility on signature number)
- CRUD operations supported **without downtime**
- Conformance to the **NFV paradigm** (scalable by adding virtual processor instances) → may be offered as a service in cloud infrastructures

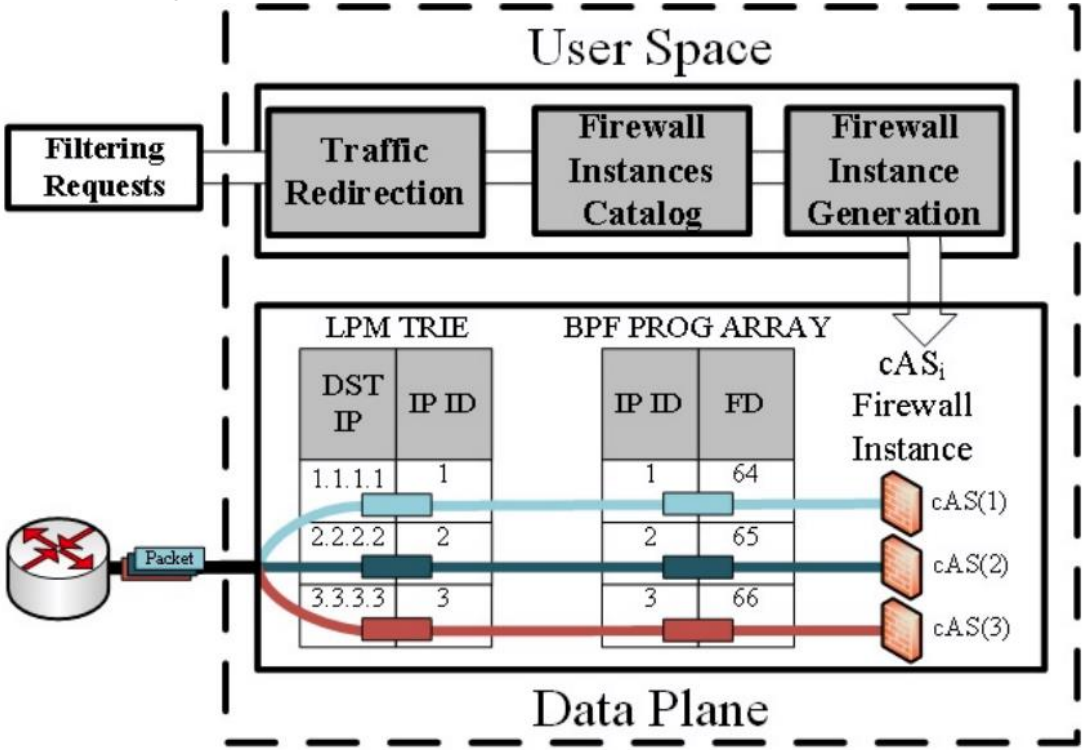
DDoS Mitigation App (2/2)

Signatures are converted to Firewall Instances (FI's) via appropriate Jinja templates

Firewall Instances (FI's): Instantiations of filtering rules that:

- Parse packet fields to derive signatures
- Match and drop malicious packets based on if-then-else conditions
- Are indexed by unique File Descriptors (FD's)

Appropriate types of eBPF maps (LPM TRIE, BPF PROG ARRAY) are used to "route" ingress packets to the appropriate Firewall Instance based on the packet destination IP



Collaboration Manager (CM)

Based on extensions of multi-domain environment legacy tools, i.e. BGP signaling

Handles filtering requests for/from collaborators

CM's involve a **BGP speaker** supporting the **Content-URI** address family

→ Advertisement of specialized BGP Updates with URI's to requested application-specific filtering rules (payload signatures, e.g. type ANY DNS requests)

Coordinates the cross-silo FL training process among collaborators without exchanging sensitive data

A **message broker** (RabbitMQ) that:

- Authenticates collaborators
- Enables inter-collaborator agreements (e.g. FL training termination)
- Retrieves local weights and reliably delivers them to the FM

Experimental Evaluation

- Comparison of signature classification accuracy for:

  - Individual detection models of participants (no collaboration)
  - Our proposed privacy-aware Federated Learning detection model

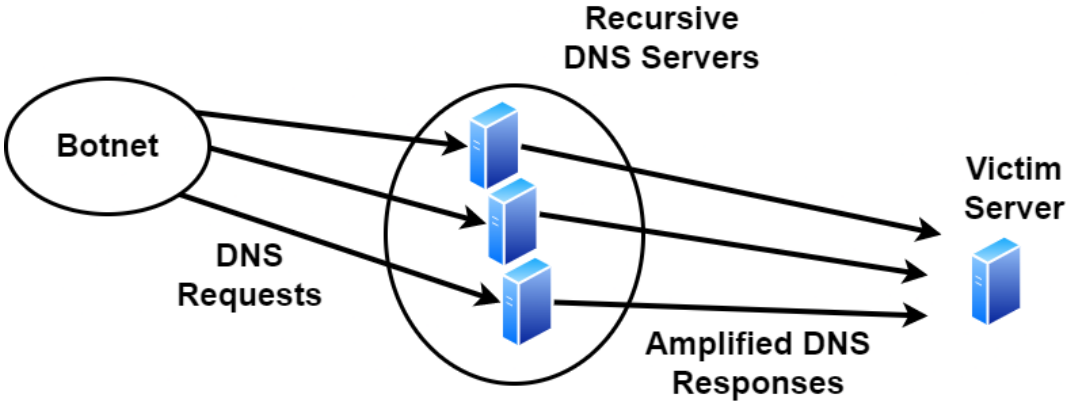
- Evaluation of packet processing performance for our **data plane programmable firewalls**, considering multiple victims/attacks

Based on emulation of attack scenarios performed within our laboratory testbed Virtual Machines (VM's)

Use Case: DNS Amplification Attacks

Target: Assessing the benefits of Federated Learning for collaborative DDoS protection, hence consider a single attack vector

One of the most common and devastating DDoS attack vectors



Selected features were based on previous work

Packet Fields (Features)		
ip.length	*dns.flags.checkdisable*	*dns.count.answers*
udp.length	*dns.flags.authoritative*	*dns.count.auth_rr*
dns.qry.name	*dns.flags.truncated*	*dns.count.add_rr*
dns.qry.type	*dns.flags.recdesired*	*dns.flags.recavail*
-	*dns.flags.authenticated*	-

- Other attack vectors (e.g. TCP SYN attacks) could be considered by appropriately selecting machine learning model packet features

Datasets – DN4 Traffic Traces

Baseline (Benign) Traffic

- 10G link: WIDE-G backbone and DIX-IE Internet Exchange (Japan)
- Aggregation of packets by destination AS using public BGP data (~**40,000** BGP messages/day)
- AS's sorted in descending order based on total received packets
- Traffic of 7 highest ranking AS's: B_1, B_2, …, B_7

Malicious (Attack) Traffic

- Booters dataset: Attack traffic captured by researchers studying DDoS-for-hire services
- 7 DNS amplification attacks (~**5** Gbytes) denoted as A_1, A_2, …, A_7
- A_1, A_2, A_3, A_6 and A_7 based on type ANY DNS responses, A_4 and A_5 based on type A DNS responses

Publicly available datasets

Signature Classification Accuracy (1/2)

Target: Investigate if our FL detection schema increases signature classification accuracy compared to individual models of collaborators

- We assumed **7 collaborating AS's**, each denoted as **cAS(i)**, i = 1 .. 7
- Each **cAS(i)** has access to its private traffic sets:
  - The attack traffic set $\mathbf{A_i}$
  - The benign traffic set $\mathbf{B_i}$
- **Binary classification** for traffic signatures (attack/benign)
- Local Machine Learning model training details:

- MLP's: 13 input neurons (# features), 27 hidden neurons[1]
- Weight optimizer (Adam), Hyperparameter optimization (Grid Search)
- We ignored inconsequential packet fields whose values were:
  (i) identical in the attack and benign traces
  (ii) randomly generated packet field values, e.g. DNS ID

1. C. Siaterlis and V. Maglaris, "Detecting Incoming and Outgoing DDoS Attacks at the Edge using a Single Set of Network Characteristics", Symposium on Computers and Communications, June 2005, pp. 469–475.

Signature Classification Accuracy (2/2)

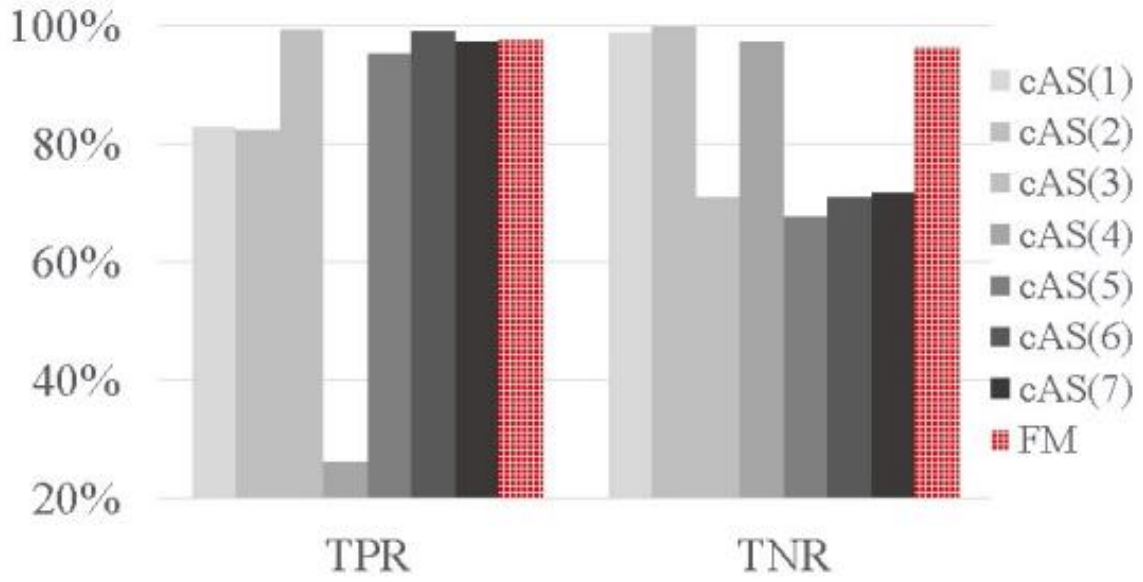
Training Datasets

- cAS(i) model is trained on: (i) Attack traffic A_i and (ii) Benign traffic B_i
- Federated Model (FM): Trained via Fed. Averaging in fewer than 40 iterations of cAS(i) models

Testing Dataset: All attack and benign traffic sets A_i and B_i

True Positive Rate (TPR):
Percentage of attack packets correctly classified as malicious

True Negative Rate (TNR):
Percentage of benign packets correctly classified as benign



Our FL scheme (red bar) enables all collaborating parties (grayscale bars) to identify benign and attack packets that as individuals might misclassify them

Mitigation Performance (1/2)

Target: Assess packet filtering performance of our **programmable firewalls** considering CPU scalability capabilities and the number of supported Firewall Instances (FI's)

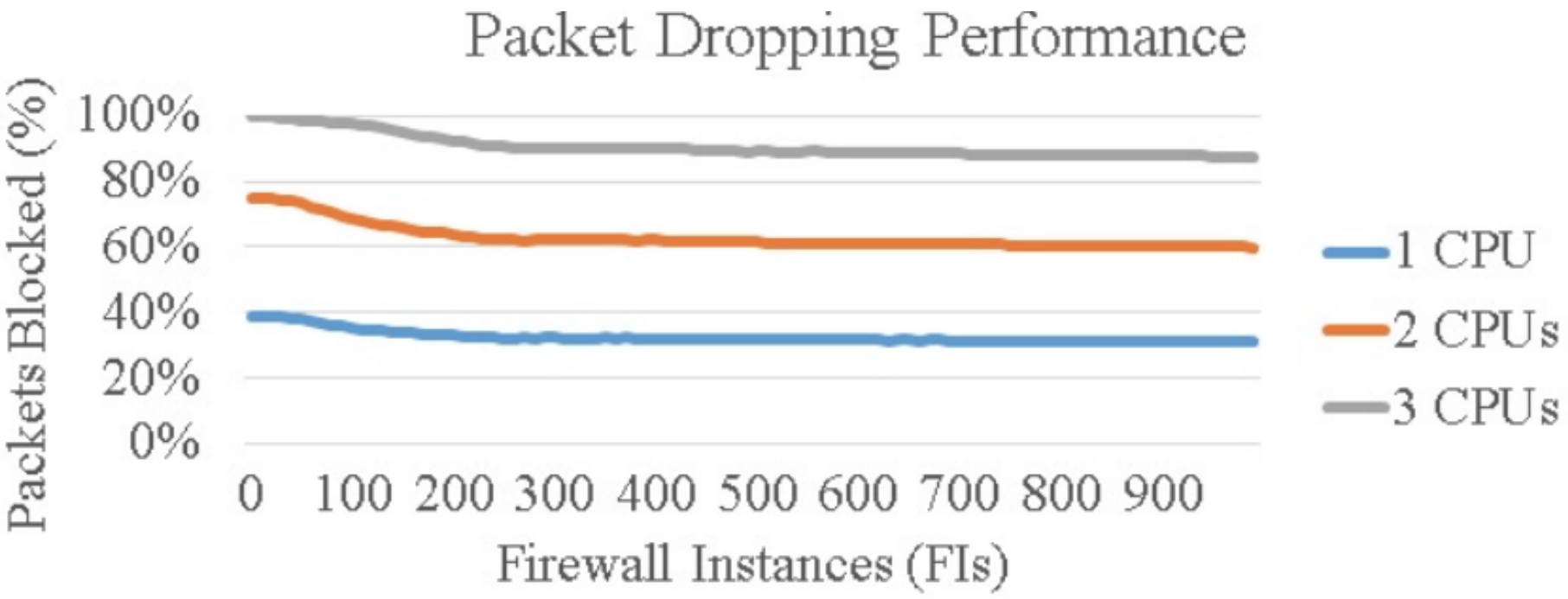
Scenario:

- Concurrent DNS amplification attacks ranging from 10 to 1000
- Attacks target different collaborators, i.e. dst IP's (unique FI per IP)
- Total attack throughput: 10 million packets per second
  - **Attacker VM:** High-speed packet generation (PF_RING ZC)
  - **Mitigation VM:** XDP-enabled *Netronome* SmartNICs

Evaluation of our programmable firewall scalability in terms of:

- The number of deployed FI's implemented as a VNF within a VM
- The number of available vCPU cores (1, 2 or 3)

Mitigation Performance (2/2)



- Processing performance scales almost linearly with the core number
- Increasing FI number decreases firewall processing rate
- Our approach can handle successfully up to 1000 concurrent attacks targeting an equal number of collaborators

Conclusions & Future Work

- Our FL model enabled collaborators to classify benign/attack packets with increased accuracy compared to individual models
- During massive attacks our schema enabled victims to raise filtering requests on collaborating domains to promptly block them
- Our programmable firewall successfully mitigated high-rate attacks, proving scalable for evolving cyber infrastructures

Future Work:

- Detection of multiple attack vectors using Multi-Task Learning
- Explore federated tree-based algorithms instead of MLP's
- Trust-based schemes to improve FL performance and security
- Compare XDP-based mitigation mechanism with other data plane programmability techniques, e.g. P4 and DPDK
- Distribute operation of single point of failure components by using distributed ledger technologies, e.g. Blockchain

Πρόσφατες Σχετικές Δημοσιεύσεις Ερευνητικής Ομάδας

NetMan @ netmode.ece.ntua.gr

- "**Orchestrating DDoS Mitigation via Blockchain-based Network Provider Collaborations**", **Adam Pavlidis**, **Marinos Dimolianis**, **Kostas Giotis**, **Loukas Anagnostou**, **Nikolaos Kostopoulos**, **Theocharis Tsigkritis**, **Ilias Kotivas**, **Dimitrios Kalogeras** & **Vasilis Maglaris**, *The Knowledge Engineering Review*, Volume 35, pp. 1-17, April 2020

- "**Signature-Based Traffic Classification and Mitigation for DDoS Attacks Using Programmable Network Data Planes**", **Marinos Dimolianis**, **Adam Pavlidis** & **Vasilis Maglaris**, *IEEE Access*, Volume 9, 2021, pp. 113061-113076

- "**DDoS Attack Detection via Privacy-aware Federated Learning in Multi-domain Cyber Infrastructures**", **Marinos Dimolianis**, **Dimitrios Kalogeras**, **Nikos Kostopoulos** & **Vasilis Maglaris**, *11th IEEE International Conference on Cloud Networking* (CloudNet), Paris, France, November 2022, pp. 118-125

- "**SHAP Interpretations of Tree and Neural Network DNS Classifiers for Analyzing DGA Family Characteristics**", **Nikos Kostopoulos**, **Dimitris Kalogeras**, **Dimitris Pantazatos**, **Mary Grammatikou** & **Vasilis Maglaris**, *IEEE Access*, Volume 11, 2023, pp. 61144-61160