

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

**ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

**Σεληνιωτάκη Αλέκα**

**Μεταπτυχιακή Φοιτήτρια**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Βασίλης Χριστοφίδης

**Τετάρτη, 18 Μαρτίου 2015, 11:30**

**Αίθουσα K206, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**" Παρακολούθηση συσχετίσεων μεταξύ ροών δεδομένων σε συστήματα  
επεξεργασίας δεδομένων με αβεβαιότητα"**

#### **ΠΕΡΙΛΗΨΗ**

Σε πολλές βιομηχανικές εφαρμογές η παρακολούθηση υποδομών μεγάλης κλίμακας είναι υψηλής σημασίας, προκειμένου να παρέχονται οι κατάλληλες ειδοποιήσεις για μη αναμενόμενη συμπεριφορά. Αυτός είναι και ο λόγος που αυξήθηκε η τάση για την ανάπτυξη δικτύων αισθητήρων μεγάλης κλίμακας. Ωστόσο, αυτή η επεξεργασία από τα δίκτυα αισθητήρων καταλήγει στο χειρισμό μεγάλου όγκου δεδομένων πληροφορίας με χαμηλή ποιότητα. Η αναξιοπιστία στα δεδομένα επεξεργασίας, αποτελεί έναν ανασταλτικό παράγοντα στα συστήματα διαχείρισης δεδομένων πραγματικού χρόνου των οποίων η ζήτηση είναι αυξημένη. Σε αυτήν την εργασία προτείνουμε ένα σύστημα επεξεργασίας δεδομένων χωρίς αβεβαιότητα, το οποίο είναι ικανό να παρακολουθεί τη συμπεριφορά και τις αλληλεπιδράσεις μεταξύ μεγάλων και ετερογενών ροών δεδομένων από αισθητήρες, σε πραγματικό χρόνο. Η προσέγγισή μας χρησιμοποιεί μία

συνάρτηση ομοιότητας, αντί τον τυπικό συντελεστή συσχέτισης , έτσι ώστε να πραγματοποιείται η παρακολούθηση στην εξέλιξη δυναμικών φαινομένων για την έγκαιρη ανακοίνωση σημαντικών ειδοποιήσεων, καθώς και για την εξασφάλιση της εγκυρότητας για ανίχνευση σημαντικών γεγονότων. Η πειραματική μας αξιολόγηση σε ένα σύνολο πραγματικών δεδομένων τα οποία καταγράφονται από αισθητήρες σε μία βιομηχανική μονάδα αφαλάτωσης νερού, εμφανίζει αυξημένη απόδοση όσον αφορά στην επίτευξη μειωμένων χρόνων απόκρισης του συστήματος. Επιπλέον, παρουσιάζει μεγάλη ακρίβεια στην ανίχνευση υψηλά συσχετιζόμενων ροών δεδομένων, σε αντίθεση με διαφορετικές τεχνικές επεξεργασίας ροών δεδομένων.

## **“Stream Correlation Monitoring for Uncertainty-Aware Data Processing Systems”**

### **ABSTRACT**

In several industrial applications, monitoring large-scale infrastructures in order to provide notifications for abnormal behavior is of high significance. For this purpose, the deployment of large-scale sensor networks is the current trend. However, this results in handling vast amounts of low-level, and often unreliable, data, while an efficient and real-time data manipulation is a strong demand. In this thesis, we propose an uncertainty-aware data management system capable of monitoring interrelations between large and heterogeneous sensor data streams in real-time. To this end, an efficient similarity function is employed instead of the typical correlation coefficient to monitor dynamic phenomena for timely alerting notifications, and to guarantee the validity of detected extreme events. Experimental evaluation with a set of real data recorded by distinct sensors in an industrial water desalination plant reveals a superior performance of our proposed approach in terms of achieving significantly reduced execution times, along with increased accuracy in detecting extreme events and highly correlated pairs of sensor data streams, when compared with state-of-the-art data stream processing techniques.