Energy Efficient In-Network Phase RFID Data Filtering Scheme[⋆]

Dong-Sub Kim, Ali Kashif, Xue Ming,
Jung-Hwan Kim, and Myong-Soon Park[⋆⋆]

Department of Computer and Radio Communications Engineering,
Korea University, Seoul, 136-713, Korea
{gastone, ali, xueming, glorifiedjx, myongsp}@ilab.korea.ac.kr

Abstract. One of the limitations that prevent proliferation of RFID technology is redundant data transmission within the network usually caused by unreliability of readers and duplicate readings generated by adjacent readers. Such redundancies unnecessarily consume resources of network and depreciate the performance of RFID installation. In this paper, we propose a CLIF, an energy-efficient filtering scheme that detects the in-network redundant data and eliminates it. The simulation results show that the CLIF significantly reduces the number of comparisons required for detecting duplicates while it achieves relatively high duplicate data elimination ratio considering the location of reader. Consequently, CLIF reduces the considerable amount of transmission within the network.

1 Introduction

RFID (Radio Frequency Identification) technology provides identification of each object using radio frequency signals. Unlike conventional techniques such as barcodes, RFID does not have a limitation of 'line of sight'. Recent development of RFID technology is boosting its applications' rapidly while RFID devices e.g. tag are becoming increasingly smaller and cheaper. The Auto-ID [9] Center has proposed methods which could lower the cost per RFID chip to approx. 5 U.S. cents. Consequently many industries such as supply chain management, inventory and military applications showed a great thirst of RFID technology.

1.1 Applications and Challenges

RFID technology is facilitating the business industry in various manners. It has been gradually adopting in many application areas. In applications such as military equipment management, large-scaled inventory systems, RFID installations have to be deployed in a wide area. Covering these large areas with RFID technology requires system infrastructures such as wired facilities or APs (Access Points). However in case of military equipment management systems, it will be

---

[⋆] This work was supported by the Second Brain Korea 21 Project.

[⋆⋆] Corresponding Author.

F.E. Sandnes et al. (Eds.): UIC 2008, LNCS 5061, pp. 311–322, 2008.

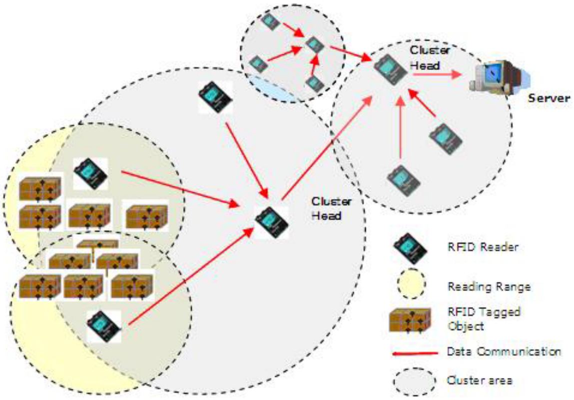


Fig. 1. Example of RFID System with Mobile RFID Readers

hard to implement infrastructures while area is not defined in it or presence of enemies and their movement could be agile and fast. Therefore, implementing RFID systems without a define infrastructure might be beneficial for this application.

Implementation of infrastructure-less RFID systems can be facilitated by using WSN (Wireless Sensor Network) along with mobile RFID readers. Mobile RFID reader is the handheld device with RFID antennas and has a capability of wireless communication. Fig. 1 represents our RFID system architecture showing the infrastructure-less nature of the system. While working with WSN topology and considering its characteristics, energy consumption is a critical point need to be considered.

In the RFID, however, readers being unreliable in nature manipulate duplicate readings. The observed read rate (i.e., number of tags read to the actual number of tags in a reader vicinity) in real-world RFID installation is 60-70% [8]; in other words, each reader produces a lot of duplicate readings for achieving accuracy [11]. Moreover readers are deployed densely so they usually have overlapping read ranges which also cause duplication in the network.

Scarce resources of wireless sensor networks put constraints; on the network lifetime energy being the most critical one. The duplicate readings cause network congestion and transmission delay since massive RFID data transmission with limited bandwidth between readers in wireless sensor network [1]. Such redundant transmission consumes a lot of energy. Filtering of duplicate data can save a lot of energy of the network and it enhance the lifetime of the network.

1.2 RFID Background

RFID installations usually consist of three components: Reader, Tags and Antenna. RFID reader communicates with tags using antenna through RF signal to produce the tag ID list. Antenna has pre-determined reading range and it can

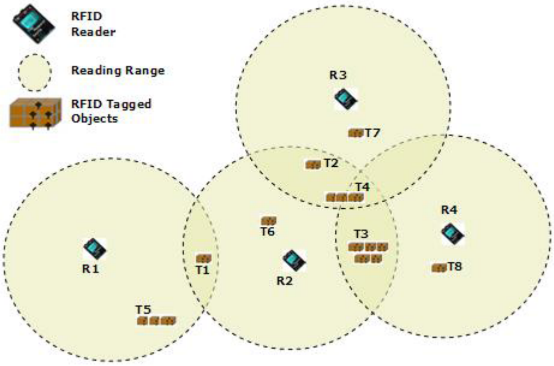


Fig. 2. Overlapped Reading Area of RFID Readers

read the information written on tags within its reading range. Unreliable nature of readers and overlapped reading vicinity among readers can cause duplicate data generation. [2] Defines data redundancy in RFID systems which can be divided into two parts: data level and reader level.

- *Redundancy at Data Level*
  When reader interrogates a tag multiple times for accuracy or reads multiple copies of tags attached on one object for reducing missing rate, reader generates duplicate readings.
- *Redundancy at Reader Level*
  Usually numerous readers are installed densely to cover whole subjected area and these readers may have overlapped reading vicinity with their neighbor readers. When the two or more than two neighboring readers interrogate tags in the overlapped area, duplicate readings are generated as shown in Fig. 4. For instance, readers R1, R2, R3 and R4 are redundant with each other so they generate duplicate readings.

1.3 Motivation

Cluster with dense deployment of readers increases huge amount of duplicated data. Output stream of a reader usually contains many duplicate values which are also described above as redundancy at data level. In a cluster, every reader sends the data to its cluster head and due to overlapping of readers more duplication is being added to the input stream of cluster head buffer which defined as redundancy at reader level. These duplicate data values can cause network congestion and increases transmission delay which need to be removed within the network to save energy and to enhance network lifetime. Therefore; an energy efficient filtering scheme is being proposed in this paper that outclasses the previous in-network duplicate data elimination approaches by considering the readers location and is a remarkable enhancement in the field of energy efficiency in RFID systems along with WSN for RFID applications.

1.4 Contributions

In this paper, an energy-efficient RFID data filtering scheme is being proposed named as CLIF (Cluster-based In-network phase Filtering scheme). Contrast to the filtering approaches at BS; the CLIF filters duplicate data within the network. For better understanding of the in-network redundancy, we divide data redundancy in following parts

- *Intra-Cluster Redundancy*
  When the member nodes want to send the data to the B/S, they send the data to the CH first and CH filter data to remove duplications. After filtering they send the data via a route to the base station.

  Redundancy occurred within the premises of a cluster is called intra-cluster redundancy. It includes:
  - Redundancy at data level
  - Redundancy at reader level
- *Inter-Cluster Redundancy*
  Densely deployed readers have overlapping areas that are covered by neighboring readers and after forming clusters; readers of one cluster might have overlapping areas with readers of neighboring clusters is called redundancy at cluster level. This kind of redundancy have not considered yet by any of previous in-network filtering approaches.

CLIF provides solution to both intra and inter cluster duplication and reduces considerable amount of network congestion; caused by duplicate data values; hence the network delay. The main challenge for RFID data filtering in cluster-based networks is elimination of redundant data generated by adjacent clusters. This approach exploits the reader's location and eliminates the duplicate data. Simulation shows that CLIF requires less number of comparisons for detecting duplicates and possess relatively high duplicate data elimination ratio compared to INPFM [1]. We also showed that CLIF's processing time is much shorter than the previous approaches.

Rest of paper is organized as; in section 2 we will give the related work of filtering algorithms. Section 3 is divided into two sub-sections; 3.1 talks about the detection of duplicates and 3.2 includes the filtering algorithm in detail. In the next section we showed the simulation results of our approach and lastly the conclusion of this paper is given in section 4.

2 Related Works

Most of literature [4], [5], [6] concentrates the data filtering at base station or sink node. In these approaches RFID readers read tags, and then the readings are sent to the B/S via a route and B/S filter the data. These approaches are quite heavy to adopt within the network as our reader have very limited resources and processing these fleshy algorithms on the reader can reduce the lifetime of device. Moreover; they being executed at BS can not reduce the amount of in-network transmission. One famous algorithm running at BS is Shawn R. Jeffery, et al proposed SMURF sliding window for RFID filter [5] on the RFID middleware or BS. They compute an adaptive window size to detect duplications and then RFID data filtering is done with a sliding window. They assume that adjacent readers can read a same object, so the proposed filter eliminates the data which has a same EPC value in a sliding window. But the window size is a critical issue in SMURF and higher the window size will result more reliable data.

[7] Suggests a simple solution in case of redundant readers. It chooses some of the readers to be active and rest may be turned off. But the problem of finding the optimal solution for the redundant reader elimination i.e. which reader to be turned on; is NP-hard problem [10]. [1] Proposed an In-Network Phased Filtering Mechanism (INPFM). This mechanism works on every node. However, the nodes closer to sink will consume a lot of energy for relaying the data of downstream clusters to the BS; so they filter RFID data when 'k' value is less than the number of hops between the node itself and the BS. Otherwise, they transmit the data to the next intermediary node directly. However, filtering depends on the value of 'k'. if k is high then huge number of CH's will be involved in filtering. Since we may know the probability of duplication by considering the reader's location, we can apply filtering procedure only on few nodes. [1] Only remove reader level duplication and it also have unnecessary comparisons to filter the data which can be avoided to save a lot of energy and overload of the network.

3 CLIF (Cluster-Based In-Network Phase Filtering Scheme)

For our network model we use the cluster-based wireless sensor network into RFID installation using RFID mobile reader where each reader acts like a node and are in the shape of clusters. We assume the sensor network protocols to be followed for our approach. To detect and eliminate duplicate readings, we proposed an energy efficient filtering approach named as *CLIF*, which can be divided into two phases as given below

- Phase 1: It includes duplicate detection scheme in case of inter and intra-cluster duplication. Explanation is given in Section 3.1
- Phase 2: An efficient in-network RFID data filtering method called *RSF* (Reverse Search Filtering) is presented in Section 3.2

3.1 Duplicate Detection Scheme

In the cluster-based RFID network, there exist two kinds of duplicate data such as Intra-cluster Duplication (*IRD*) and Inter-cluster Duplication (*IED*). In *IRD*, every member sends the data to the CH (Cluster Head) and CH filters the data. Through this procedure, CH can eliminate the redundancy of both data level and reader level.

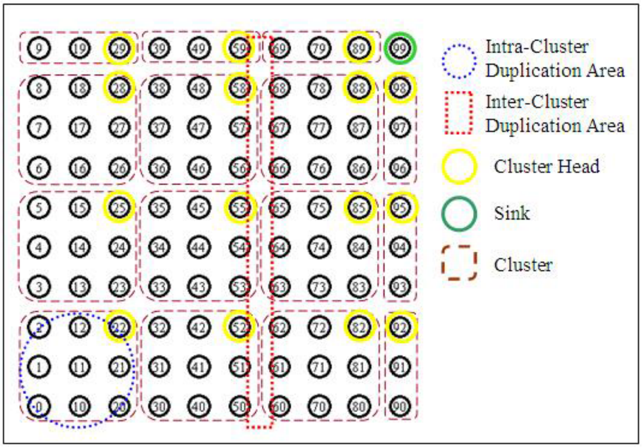


Fig. 3. Duplicate Readings in Cluster-Based RFID System

Cluster-based network generates one more kind of duplication, *IED* where readers of neighboring clusters can be redundant. In dense network topology there could be immense redundancy at cluster level as shown in Fig. 3. Such kind of duplication however, can not be removed in the intra-cluster filtering phase as the CH can only eliminates the duplicate reading of its own cluster. Therefore *IED* can be detected and eliminated by an upstream cluster while forwarding data to base station.

The main challenge of *IED* filtering is to find duplicate data among clusters efficiently. For the above reason, we consider the readers' location. To eliminate the duplicate data between inter-cluster areas, we select an appropriate CH which has a chance to be redundant. CH in the route path to the base station filters duplicate data so that they can reduce the number of transmission data packets. *CLIF* addresses both these intra and inter-cluster filtering separately as given below.

- *Intra-Cluster Filtering Step*
  When the member nodes want to send the data to the B/S, they send the data to the CH first and CH filter data to remove duplications. After filtering they send the data via a route to the base station.
- *Inter-Cluster Filtering Step*
  If CH receives the data from other cluster, they check the source's cluster id and if it is neighboring cluster then they process an inter-cluster filtering step.

Duplicate Decision Phase:
Node which receives data checks whether I am CH. If it is not, it forwards data to its own cluster head. If it is CH, it checks the source cluster id whether it destined to neighbor cluster or to myself. Address of a node can be used to

generate cluster id. This algorithm facilitates us to select the CH which has probability of being duplicated.

Algorithm 1. **Filtering Step Decision.**

```
filtering step decision (params: incoming data)
 loop for incoming data
  if i am a CH then
    $s\_cid \leftarrow$ cluster id of incoming data
    if $s\_cid$ is same as my cluster ID then
      filtering step
    else
      search $s\_cid$ from my neighbor cluster IDs
      if found then
        filtering step
      else
        send the data through the route
      end if
    end if
  else
    send the data to the CH
  end if
 end loop
```

3.2 An Efficient RFID Data Filtering Method

Firstly we will define the duplication and data model based on EPC [11]. When an RFID reader read a tag, the following triplet is read.

RFID Triplet = EPC Code, Reader ID, Timestamp

When the reader reads two triplets, duplication occurs if the following three conditions described in the definition of duplication are satisfied:

- *Definition of Duplication*
  Triplet A = E_A, R_A, T_A
  Triplet B = E_B, R_B, T_B
  1) The value of E_A is equals to the value of E_B
  2) The R_A and R_B are same or neighboring readers.
  3) Difference between T_A and T_B is lesser than constant T

RFID Reader integrated with a sensor node will eliminate duplicate tag data. One of the conditions of duplication is the same EPC value. EPC consists of four parts - a header, a general manager number, an object class, and a serial number. In this EPC structure, usually objects have different serial number. Although different object can have same serial number, but this situation occur rarely in the domain of RFID applications.

Algorithm 2: **Filtering Method.**

```
tagdata
{ epc,
  reader_id,
  timestamp
}
Filtering (params: tagdata)
{
  hash_key hk1 = hash(tagdata.epc.serial number)
  search hk1 from hash table
  if not found then
    insert (hk1, tagdata) to the hash table
  else if found then
    if rfid_tagdata.timestamp -
        hk1.tagdata.timestamp < time_interval then
      if is_neighbor() then
        compare tagdata.epc.others and
          hk1.tagdata.epc.others
        if same then
          drop rfidobject as duplicate data
        end if
      end if
    end if
    update tagdata to hash_table using hk1
  end if
}
```

RSF (Reverse Search Filtering) is proposed to filter the data efficiently so we can send the data quickly to the upstream. [1] Discuss the similar kind of filtering method and it compares serial part of EPC first, and then it compares other part of EPC, reader ID and timestamp respectively. However, comparing other part of EPC after comparing the serial number is in-efficient. RSF compares reader ID or timestamp and it is more efficient because the probability of having different other part of EPC is lesser than the probability of having different reader ID or timestamp in EPCs which have a same serial part.

4 Simulation Results

In this section, we evaluate the performance of our proposed algorithm. We calculate the number of raw data, *IRD* and *IED* in Section 4.2. We evaluate the *NC* and *DER* of *CLIF* with INPFM [1].

- *Definition 1.The Number of Comparison (NC)*
  We define the NC as the number of comparison for examining duplication of

data through in-network transmitters. For example, when data A is transmitted from a specific node to sink and if data comparison to examine the presence of duplication is performed at some CHs, say 3 times, during transmission, the NC value is 3.

- *Definition 2. Duplicate data Elimination Ratio (DER)*
  The *DER* represents the ratio of duplicate data elimination.

4.1 Simulation Settings

In this part, we considered grid network topology and simulation environment is given below in table 1. According to the table 1, duplicate ratio between two neighboring nodes is 10%, which means if a node has 2 neighbors, they produce 4 duplicate readings. As considered a cluster-based network and the number of clusters vary from 16 to 225. Each cluster has 9 members and right node at the top in the cluster is selected as a Cluster Head (CH). The sink node exists at the rightmost side on the top of the network as shown in Fig.3.

4.2 Relation between the Number of Cluster and Duplicate Data

As discussed the relation among the raw data, *IRD* and *IED* by changing the number of clusters. Let's suppose the number of node in the square area is M2 and the amount of data generated from one node is n. And we also suppose the number of cluster is C2 and the duplication ratio between one neighbor and node itself is d_p. Then, we can calculate the both identical data and data redundancy.

- *Raw Data*
  $= (nM^2)\ C^2$
- *Duplicate data in Intra-cluster area (IRD)*
  $= 2C^2\ nd_pM(M\text{-}1)$
- *Duplicate data in Inter-cluster area (IED)*
  $= 2nd_pM\ C(C\text{-}1)$

Assigning our simulation parameters to the equations above, we can get the following graph shown in Fig. 4.

Table 1. Parameters used in simulation

Item	Value
Number of node	144 ˜ 2025
Reading range	5m
Distance between nodes	8m
Number of Tags in each reading area	20
Duplication ratio per neighbor node	10%
Number of members in each cluster	9
Number of cluster	16 ˜ 255

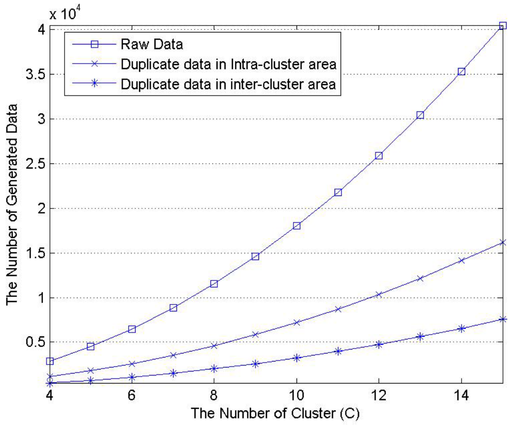


Fig. 4. Ratio of Raw data and *IRD*, *IED* data with varying number of Clusters

The Number of Nodes (M^2) = 9
Duplication probability (d_p) = 0.1
The Number of reading in one node (n) = 20

In the Fig. 4, as the number of clusters increases, *IRD* and *IED* also increase. In the next section, we vary the number of clusters from 42 to 152 to observe the NC and the *DER* compared to INPFM.

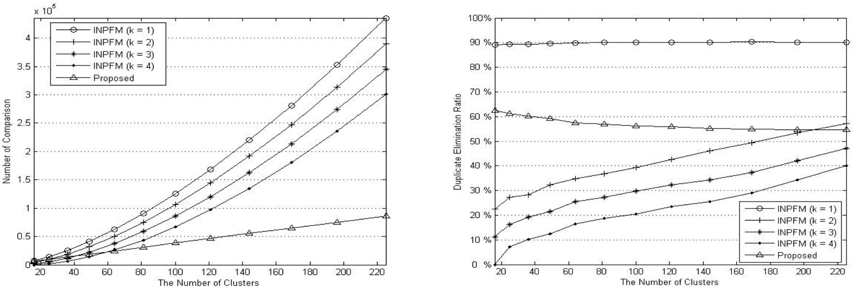
4.3 Evaluation of *NC* and *DER*

Fig.5 (a) shows that *CLIF* has less NC with varying number of clusters compared to INPFM. We vary the value of 'k' from 1 to 4. In this simulation as the number of clusters increases, the NC will also increase. In other words, if the number of hops between nodes and sink increase, the efficiency of filtering at INPFM decreases. For clusters size 25 with $k = 4$; *NC* of *CLIF* is 71% less than that of INPFM.

The less NC does not mean the existence of *DER*. Fig. 5(b) shows the *DER* with varying the number of clusters. Even though NC of INPFM is 4 times greater than that of *CLIF*, it will have lower *DER*; except the value k is 1. Overall, *CLIF* can be a good trade-off between *NC* and *DER*, as it guarantees of more than 50% *DER* and it requires lesser *NC* compare to INPFM.

4.4 Filtering Time

In order to evaluate performance of our *RSF* algorithm, we simulated one reader during 10ms and configured the constant T on condition of duplication to 1ms. The processing time at *CLIF* is approximately 8 10% faster than the processing time at INPFM as shown in Fig. 6. Furthermore, when the data exceed more than 20,000, the processing time becomes lesser than 40msec. On the other



(a) *NC* in varying number of clusters (b) *DER* in varying number of clusters

Fig. 5. *NC* and *DER* in varying number of clusters

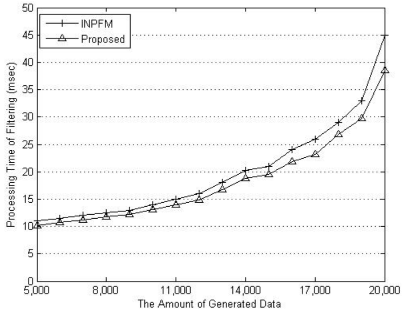


Fig. 6. Comparing Processing Time between INPFM and Proposed Filter

hand INPFM takes 45msec. Hence, we can guarantee that *CLIF* outperform the INPFM and it process the duplicate data much faster than INPFM and decrease the transmission delay.

5 Conclusion

Transmission of redundant data packets consumes immense amount of energy and cause the delay in the network. By eliminating duplicate data we can save precious resources of the network. In this paper, we defined the duplication in cluster based networks. Duplication of data in Cluster based networks is distinguished in two categories (1) Intra-cluster duplication (*IRD*) and (2) Inter-cluster duplication (*IED*). *IED* occurs when readers of neighboring clusters overlap with each other; is first time considered in this paper. Moreover we proposed energy efficient duplicate data filtering scheme *CLIF* (Cluster-based In-network phase Filtering scheme). *CLIF* provide efficient solution to both *IRD* and *IED*. *CLIF* consist of two steps. First one is the detection of duplicate data within the network and secondly it includes an algorithm names as *RSF* (Reverse Search Filtering) which expeditiously remove the duplicate values. We evaluated our approach by appropriate simulation results showing that even though *CLIF* has lesser NC required for detecting duplicate data, the *DER* is still higher than others in-network data filtering schemes.

References

[1] Choi, W., Park, M.-S.: In-Network Phased Filtering Mechanism for a Large-Scale RFID Inventory Application. In: The 4th International Conference on IT & Application (ICITA 2007), Harbin, China, January 15-16, vol. 2, pp. 401–405 (2007)

[2] Derakhshan, R., Orlowska, M., Li, X.: RFID Data Management: Challenges and Opportunities. In: IEEE First International Conference on RFID, March 27-28 (2007)

[3] Bohn, J., Mattern, F.: Super-Distributed RFID Tag Infrastructers, TR, Institute of Pervasive Computing, ETH (2004)

[4] Wang, F., Liu, P.: Temporal Management of RFID Data. In: The Proceedings of the 31st Very Large Data Bases Conference (VLDB 2005), August 2005, pp. 1128–1139 (2005)

[5] Jeffery, S.R., Garofalakis, M., Franklin, M.J.: Adaptive Cleaning for RFID Data Streams. In: The proceedings of the 32nd International Conference on Very Large Data Bases (VLDB 2006), September 12-15, pp. 163–174 (2006)

[6] Bai, Y., Wang, F., Liu, P.: Efficiently Filtering RFID Data Streams. In: The First International VLDB Workshop on Clean Databases (CleanDB), September 11 (2006)

[7] Carbunar, B., Ramanathan, M., Koyuturk, M., Hoffmann, C., Grama, A.: Redundant reader elimination in RFID systems. Sensor and Ad Hoc Communications and Networks, 176–184 (2005)

[8] Chawathe, S.S., Krishnamurthy, V., Ramachandrany, S., Sarma, S.: Managing RFID Data. In: The Proceedings of the 31st Very Large Data Bases Conference (VLDB 2005), pp. 1189–1195 (August 2005)

[9] Sarma, S.E.: Towards the five-cent tag. Technical Report MIT-AUTOID-WH-006, MIT Auto-ID Center (2001), `http://www.autoidcenter.org/research/MIT-AUTOID-WH-006.pdf`

[10] Carbunar, B., Ramanathan, M., Koyuturk, M., Hoffmann, C., Grama, A.: Redundant reader elimination in RFID systems. Sensor and Ad Hoc Communications and Networks, 176–184 (2005)

[11] EPCglobal Tag Data Standard (TDS) Version 1.3 document, `http://www.epcglobalus.org/dnn_epcus/KnowledgeBase/Browse/tabid/277/DMXModule/706/Command/Core_Download/Default.aspx?EntryId=297`