A Method for Removing Duplicate Ciphertext Data for Deep Defense in Wireless Sensor Networks*

GAO Junjie[1*], YANG Fan[2]
(1.School of Computer and Information Technology, Shanxi University, Taiyuan Shanxi 030006, China;
2.Department of Computer Education, Shanxi Medical University, Taiyuan Shanxi 030001, China)

Abstract: In wireless sensor networks, duplicate ciphertext data during transmission and storage can increase network load and resource consumption, reducing network performance. Therefore, a ciphertext data deduplication method for deep defense in wireless sensor networks is proposed. Firstly, the compressed sensing method is used to collect ciphertext data in wireless sensor networks. Secondly, the ciphertext data are mapped to the feature space to obtain the features of the ciphertext data. Finally, based on the dynamic weights of feature attributes, the similarity between the features of ciphertext data is calculated, and the removal of duplicate data is achieved through similarity calculation. The simulation results show that as the data deduplication time increases, the proposed method can achieve network residual energy of over 85%, network throughput of over 3.5Mb/s, and network space compression rate of over 60%. This indicates that the proposed method has high efficiency in deduplicating ciphertext data and can achieve deep defense of wireless sensor networks.

Key words: wireless sensor network; deduplication of ciphertext data; network defense; compression perception method; similarity calculation
EEACC: 7230 **doi**: 10.3969/j.issn.1004-1699.2025.10.021

面向无线传感网络深度防御的密文数据去重方法*

高俊杰[1*], 杨 帆[2]
(1.山西大学计算机与信息技术学院, 山西 太原 030006; 2.山西医科大学计算机教学部, 山西 太原 030001)

摘 要: 无线传感网络中, 重复的密文数据在传输和存储过程中会增加网络负荷和资源消耗, 降低网络性能。因此提出了一种面向无线传感网络深度防御的密文数据去重方法, 首先利用压缩感知方法采集无线传感网络中的密文数据; 其次将密文数据映射到特征空间, 获取密文数据特征; 基于特征属性的动态权值计算密文数据特征间的相似度, 通过相似度计算实现重复数据的去除。仿真结果表明, 随着数据去重时间的增长, 所提方法的网络剩余能量可达到 85%以上, 网络吞吐量可达 3.5Mb/s 以上, 且网络空间压缩率可达 60%以上, 表明所提方法的密文数据去重效率高, 可以实现无线传感网络的深度防御。

关键词: 无线传感网络; 密文数据去重; 网络防御; 压缩感知方法; 相似度计算
中图分类号: TP309 **文献标识码**: A **文章编号**: 1004-1699(2025)10-1886-06

密文数据去重是无线传感网络安全中一个重要的问题。无线传感网络[1]遭受恶意攻击而导致的信息泄露问题频繁发生, 为防止用户的信息泄露, 需要对无线传感网络展开防御, 其中, 密文数据去重是一种有效的防御手段, 密文数据[2]去重能够通过去除冗余和无效的数据, 减少网络中的数据传输量和处理负担, 同时提高数据的准确性和安全性。因此, 对无线传感网络的密文数据去重方法展开研究具有重要意义。

文献[3]方法首先基于流行度对数据展开划分, 然后用收敛加密和对称加密分别对非流行数据的内层和外层展开处理, 并通过门限秘密共享方法完成密钥共享, 最后利用收敛加密对流行数据展开处理, 以此完成加密数据的去重, 该方法去重后的网络剩余能量较低, 数据去重效果较差。文献[4]方法提出了一种以雾为中心的方法。与雾节点共享部分哈希值(Pαv), 使用可伸缩的 Bloom 哈希表(SBHT)进行内联重复数据删除。Pαv 的哈希位存储在雾节点中。非冗余数据通过非对称 Cramer Shoup 密码系统加密上传至云存储。该方法提高存储效率, 在实时物联网医疗环境中实施可行。但由于部分哈希值存储在雾节点中, 雾节点的可用性和稳定性对该算法的有效性至关重要。因此如果雾节点出现故障或不可用, 可能会影响数据删除和上传的流程。文献[5]方法首先采取表示学习获得数据的高维特征, 其次通过降维算法将高维特征降至二维散点图, 最后利用无监督聚类算法获得重复数据并对其展开去除, 该方法去重后的网络吞吐量较低,

项目来源: 国家自然科学基金重点项目(61936012)
收稿日期: 2023-11-10 **修改日期**: 2023-12-04

重复数据检测的精度较低。

为了解决上述方法中存在的问题,提出面向无线传感网络深度防御的密文数据去重方法。

1　无线传感网络的密文数据处理

1.1　密文数据采集

在无线传感网络中,由于通信环境的复杂性和节点的重叠,会导致同一份数据被多个节点重复生成和发送。同时,密文数据中存在隐私泄露、信息安全威胁、数据篡改等问题。通过密文数据采集可以追踪和记录传感网络中每个节点所生成的密文数据,可以帮助监测和分析数据传输过程中可能存在的安全隐患,及时发现并应对潜在的安全问题,保障通信数据的完整性、保密性和可用性。压缩感知技术可以结合安全需求,对采集到的密文数据进行压缩处理,以实现对密文数据的高效传输和处理,同时也需要考虑保护数据的隐私和安全性。因此,针对无线传感网络中的密文数据,面向无线传感网络深度防御的密文数据去重方法利用压缩感知[6]方法对其展开采集。

压缩感知是一种在采样过程中能够实现采样和压缩同时进行的技术,能够利用采样结果以较低的采样率恢复稀疏信号。在密文数据去重中采用压缩感知方法,能够通过对密文数据进行压缩感知采样,获取数据的部分信息并实现去重操作,从而达到隐私保护、资源节省的效果。

压缩感知的链式拓扑结构如图 1 所示。

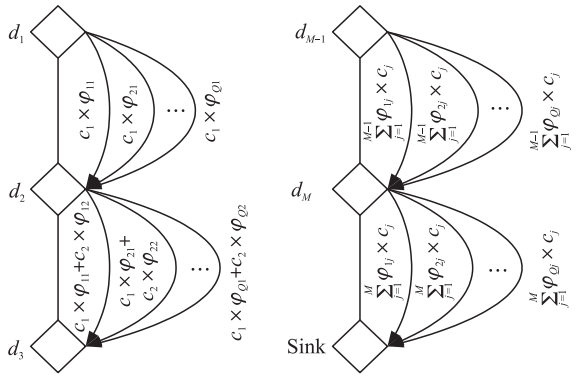


图 1　链式拓扑结构

图 1 中的无线传感网络由 M 个传感器节点与 1 个 Sink 节点构成,设定 d_i 为其中的一个传感器节点,c_j 为 d_i 的感知密文数据,i,j 均为常数项,j 代表传感器个数。利用压缩感知采集数据时,使无线传感网络的所有节点传输数据的数量相等,根据图 1 可得 $\boldsymbol{\Phi}_{Q\times M}$,其表示测量矩阵,矩阵中的元素表示加密数据的权值,第 i 行的元素表达为[$\varphi_{i1},\varphi_{i2},\varphi_{i3},\cdots,\varphi_{ij},\cdots,\varphi_{iM}$]。

由此,使用压缩感知完成密文数据采集,过程如下:

步骤 1　无线传感网络中的节点 d_1 将感知到的密文数据 c_1 与其权值 φ_{i1} 相乘,并将 $c_1\times\varphi_{i1}$ 发送给节点 d_2;

步骤 2　d_2 将感知到的密文数据 c_2 和对应的权值 φ_{i2} 相乘得到 $c_2\times\varphi_{i2}$,并将从 d_1 处接收的数据和感知数据的加权和 $c_1\times\varphi_{i1}+c_2\times\varphi_{i2}$ 发送给节点 d_3;

步骤 3　按照上述方式依次传输,同时设定无线传感网络中的节点共传送了 Q 个数据包,则 Sink 节点收集到 Q 个数据包的感知密文数据 C 可表示为:

$$C=\sum_{j=1}^{M}\varphi_{ij}\times c_j \tag{1}$$

通过密文数据采集,可以将各个传感节点收集到的密文数据进行集中存储和处理,从而实现对数据的分析和挖掘,为安全防御和决策提供支持。

1.2　密文数据的特征提取

与传统的明文数据相比,密文数据受到加密算法的保护,其内容在传输和存储过程中是不可读的,只有掌握加密算法和密钥的用户才能解密,直接对密文数据进行特征提取可能难以获取原始特征空间中的有效信息。因此,在密文数据采集后,通过从密文数据中提取具有代表性的特征,以实现对数据的分类、识别或分析。

利用主成分分析法将采集到的无线传感网络密文数据 C 映射到高维的特征空间中,为实现密文数据的规范化,将密文数据转化为 $\boldsymbol{\Gamma}(C)$ 的函数形式,此时根据密文数据的函数形式建立协方差矩阵[7-8] $\boldsymbol{V}$,定义如下:

$$\boldsymbol{V}=\sum_{j=1}^{M}\boldsymbol{\Gamma}(C_j)^{Y} \tag{2}$$

式中:Y 为常数。

通过线性变换[9]的方式对密文数据的特征值展开定义,得到映射后的特征空间中的表示 b:

$$b=\sum_{j=1}^{M}s_j\boldsymbol{V} \tag{3}$$

式中:s_j 为密文数据的线性变换系数。

设定无线传感网络密文数据在特征空间中形成的投影为 u^e,其中 e 表示密文数据特征值 b 的数量,则无线传感网络密文数据特征定义如下:

$$b_C=u^e b \tag{4}$$

特征提取可以从密文数据中提取出具有代表性的特征,而无需暴露原始数据内容,从而保护了数据的隐私和安全。

2 基于相似度的密文数据去重

由于无线传感网络通常使用加密算法对数据进行保护和保密,提取密文数据特征的过程受到加密算法的限制。加密算法通常会产生随机性和不可预测性,使密文数据的特征提取变得困难,而密文数据的去重可以在一定程度上防止信息被非授权方获知。通过基于相似度的去重方法,不需要具体的明文内容,仅使用特征进行比较,减少了信息泄露的风险。因此,设定无线传感网络密文数据特征集为$\boldsymbol{B}=\{b_{C1},b_{C2},\cdots,b_{Cn}\}$,$b_{Cn}$是第 n 个密文数据特征,特征的属性集$\boldsymbol{A}=\{a_1,a_2,\cdots,a_z\}$,$a_z$为特征的第 z 个属性,b_{Cnz}是特征 b_{Cn} 的第 z 个密文数据属性对应的值。由此展开相似度计算,准确地判断是否存在重复数据。

随机选择两个密文数据特征 b_{C1}、b_{C2},则 b_{C1} 在属性 z 上的值为 $b_{C_{1z}}$,b_{C2} 在属性 z 上的值为 $b_{C_{2z}}$。设定 $F(b_{C_{1z}},b_{C_{2z}})$ 为密文数据属性值 $b_{C_{1z}}$ 和 $b_{C_{2z}}$ 间的距离,则属性值 $b_{C_{1z}}$ 和 $b_{C_{2z}}$ 间的密文数据相似度 $\mathrm{Sim}(b_{C_{1z}},b_{C_{2z}})$ 定义如下:

$$\mathrm{Sim}(b_{C_{1z}},b_{C_{2z}})=1-\frac{F(b_{C_{1z}},b_{C_{2z}})}{\max(|b_{C_{1z}}|,|b_{C_{2z}}|)} \tag{5}$$

无线传感网络密文数据特征间的相似度是指特征的全部属性相似度[10-11]的加权平均值,因为不同的特征属性对判断重复密文数据的贡献度不同,因此需要为每一个属性根据贡献度大小设置权值。定义密文数据特征的各个属性的贡献比为 $v_1:v_2:\cdots:v_z$,且密文数据特征属性的所有权值相加为 1。设定 r_z 是属性 a_z 的权重,则 $r_z=a_zv_z$。无线传感网络密文数据特征间的相似度 $\mathrm{Sim}(b_{C1},b_{C2})$ 定义如下:

$$\mathrm{Sim}(b_{C1},b_{C2})=\sum_{i=1}^{z} r_i\mathrm{Sim}(b_{C_{1i}},b_{C_{2i}}) \tag{6}$$

在对密文数据特征属性间的相似度展开计算时,如果存在任何一个属性缺失,则该缺失属性和其他属性间的相似度 $\mathrm{Sim}(b_{C_{1z}},b_{C_{2z}})$ 视为零,这导致特征间的相似度随之下降,$\mathrm{Sim}(b_{C1},b_{C2})$ 小于设定的阈值,此时特征 b_{C1} 和 b_{C2} 存在不同,但不排除 b_{C1} 和 b_{C2} 是重复特征,此时,采用固定的属性权值无法解决缺失属性影响重复特征的问题,因此采用动态权值[12]来计算特征间相似度。

定义 γ 是密文数据特征属性 z 的有效性,则当 $\gamma=1$ 时,特征属性 z 有效;当 $\gamma=0$ 时,特征属性 z 无效。

若密文数据特征中缺失任意属性或存在某个属性无效,则该特征的所有属性对应的权重都随之发生变化,以降低属性缺失对重复数据判断产生的影响。基于有效性判断重新定义属性权重 r_z,表达式为:

$$r_z'=\gamma a_z \tag{7}$$

此时无线传感网络密文数据特征间的相似度 $\mathrm{Sim}(b_{C1},b_{C2})'$ 定义如下:

$$\mathrm{Sim}(b_{C1},b_{C2})'=\sum_{i=1}^{i} r_i'\mathrm{Sim}(b_{C_{1i}},b_{C_{2i}}) \tag{8}$$

若 $\mathrm{Sim}(b_{C1},b_{C2})=1$,表明特征 b_{C1} 和 b_{C2} 完全重复,则该特征对应的两个无线传感网络密文数据为重复数据,对其中一个数据展开删除即可。

根据上述去重流程,对全部密文数据特征展开相似度计算。当两个密文数据特征的相似度为 1 时,删除其中一个数据;反正,两个数据均保留。由此,完成密文数据去重,在一定程度上降低数据传输量,减少网络负载和能耗,从而在一定程度上提高了无线传感网络的深度防御性能。密文数据的去重流程如图 2 所示。

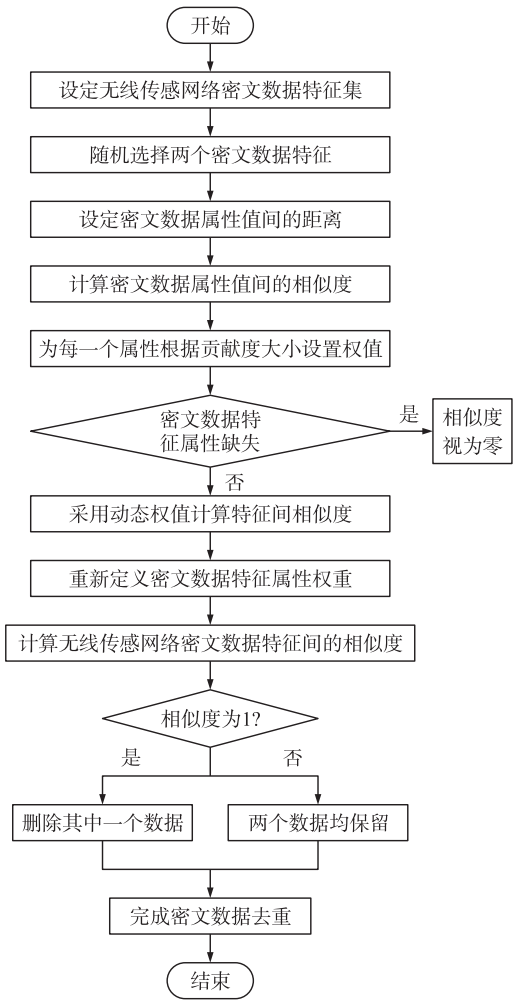


图 2 密文数据去重流程图

3 仿真与分析

3.1 仿真设置

为了验证面向无线传感网络深度防御的密文数

据去重方法的整体有效性,选取某大型实验室所需的无线传感网络系统,该系统部署了大量的传感器节点来收集环境数据并进行传输。由于网络规模庞大,数据存储和传输成本较高,且传输过程中存在较大的安全隐患,因此需要进行密文数据去重处理。该无线传感网络系统的拓扑结构如图 3 所示。

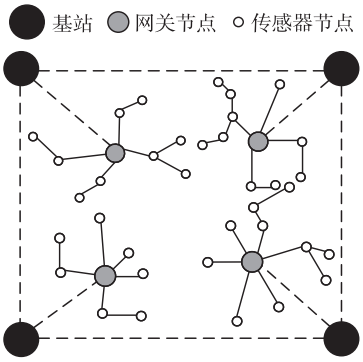


图 3　无线传感网络拓扑结构

图 3 的无线传感网络中共有 20 个无线传感器节点。其相关参数设置如表 1 所示。

表 1　无线传感网络相关参数

参数	数值
网络场景面积	200 m×200 m
MAC 层协议	IEEE 802.15.4
数据包传输速率	2 pkt/s
密文数据量	125 byte/包
仿真周期	2 000 s
节点初始化能量	15 J
节点通信半径	30 m

在无线传感网络中,使用 AES 对称加密算法对每个传感数据进行加密操作,生成相应的密文数据。包括传感器节点采集的环境参数数据、运行状态数据以及事件触发数据,被加密为密文数据进行传输。具体的密文数据相关信息如表 2 所示。

表 2　密文数据相关信息

参数	数值	
环境参数数据	温度数据	30 MB
	湿度数据	20 MB
	光照强度数据	45 MB
运行状态数据	电量数据	34 MB
	工作状态数据	52 MB
事件触发数据	特定事件数据	13 MB
	异常事件数据	22 MB

结合上述仿真设置,利用面向无线传感网络深度防御的密文数据去重方法、文献[3]基于双层加密和密钥共享的云数据去重方法和文献[4]以雾为中心的数据去重方法对无线传感网络中的密文数据展开去重处理,通过评估去重处理后的网络剩余能量、网络吞吐量和空间压缩率,说明面向无线传感网络深度防御的密文数据去重方法的有效性。

3.2　结果分析

①网络剩余能量

无线传感网络在传输数据的过程中会消耗大量的能量,网络中存在的重复数据越多,网络的剩余能量越少,生命周期越短,因此,通过无线传感网络的剩余能量可以判断网络中密文数据的去重效果。得到的网络剩余能量随数据去重时间变化的情况如图 4 所示。

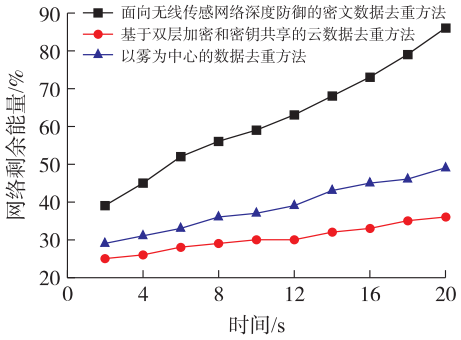


图 4　网络剩余能量

分析图 4 可知,随着数据去重时间的增长,三种方法下的无线传感网络剩余能量均有所上升,但面向无线传感网络深度防御的密文数据去重方法的网络剩余能量整体上升幅度远大于基于双层加密和密钥共享的云数据去重方法和以雾为中心的数据去重方法,且当数据去重所用时间相同时,面向无线传感网络深度防御的密文数据去重方法的网络剩余能量最大,因此,面向无线传感网络深度防御的密文数据去重方法对无线传感网络密文数据的去重效果最好。其主要原因是本文方法通过压缩感知方法采集了无线传感网络中的密文数据,有效减少了数据采集与准备时间,最大化了网络利用率,从而提高了网络剩余能量。

②吞吐量

当无线传感网络中存在的重复数据减少时,网络传输的数据总量也会减少,网络的闲置带宽就会增大,网络整体的传输吞吐量也会随之增加。现对三种方法去重后的网络吞吐量展开计算,得到的结果如图 5 所示。

由图 5 可以分析出,利用面向无线传感网络深度防御的密文数据去重方法对密文数据展开去重后的无线传感网络吞吐量较高,且吞吐量上升速度较快,而基于双层加密和密钥共享的云数据去重方法和以雾为中心的数据去重方法的吞吐量值较小且上升速

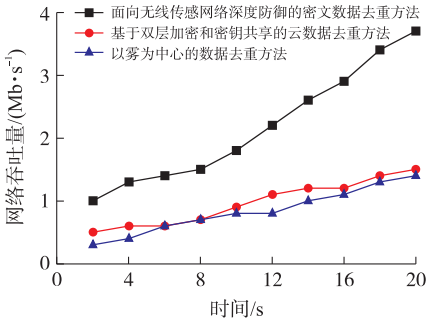


图 5 网络吞吐量

度较慢,表明在相同的时间内,面向无线传感网络深度防御的密文数据去重方法删除的重复数据数量更大,即本文方法将采集到的密文数据映射到特征空间,获取了相对完整的密文数据特征,为下一步密文数据去重做好了准备,优化了密文数据去重效果。

③空间压缩率

无线传感网络的空间压缩率是衡量网络中重复数据删除效率的重要指标,空间压缩率越高,表明该算法删除的实际重复数据越多,即算法的重复数据删除效率越高。设定 β 为无线传感网络的空间压缩率,且 $\beta\in[0,1]$,则 β 定义如下:

$$\beta=\frac{S_{\text{delete}}}{S}\times 100\% \tag{9}$$

式中:S_{delete} 表示删除重复密文数据后无线传感网络减少的空间;S 表示原始密文数据在无线传感网络中所占的空间。

为比较三种方法的重复数据删除效率,现利用式(9)对上述三种方法的空间压缩率展开计算,空间压缩率随密文数据删除时间变化的情况如图 6 所示。

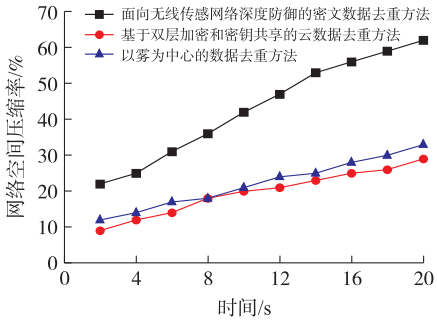


图 6 网络空间压缩率

根据图 6 可得,随着无线传感网络中密文数据去重时间的增长,三种方法去重后的网络空间压缩率均有所上升,面向无线传感网络深度防御的密文数据去重方法的压缩率整体上升幅度最大,而基于双层加密和密钥共享的云数据去重方法和以雾为中心的数据去重方法的压缩率上升幅度较小,且当数据去重的时间相同时,面向无线传感网络深度防御的密文数据去重方法的空间压缩率高于基于双层加密和密钥共享的云数据去重方法和以雾为中心的数据去重方法,表明面向无线传感网络深度防御的密文数据去重方法通过动态权值计算密文数据特征间的相似度,能快速去除无线传感网络中重复密文数据,压缩数据储存所占的空间,提高了网络空间压缩率。

④密文数据重复率

为了更加直观地体现面向无线传感网络深度防御的密文数据去重方法、文献[3]基于双层加密和密钥共享的云数据去重方法和文献[4]以雾为中心的数据去重方法对无线传感网络中的密文数据展开去重处理的效果,选择密文数据重复率作为评估指标,得到的结果如图 7 所示。

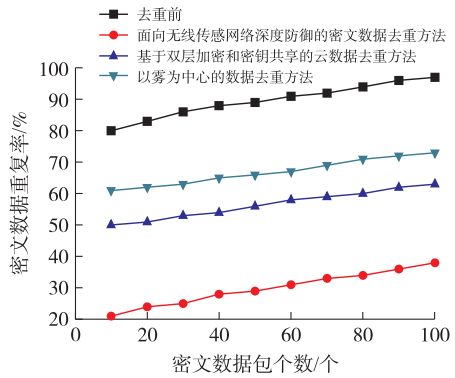


图 7 密文数据重复率

如图 7 所示,随着密文数据包个数的增加,三种方法得到的密文数据重复率均逐渐增加,其中使用面向无线传感网络深度防御的密文数据去重方法进行去重后,得到的密文数据重复率相对较低,保持在 40%以下,由此说明面向无线传感网络深度防御的密文数据去重方法更加有效。

⑤网络防御性能

结合上述实验结果,对密文数据去重对无线传感网络的防御性能影响进行实验,进一步说明对无线传感网络中的密文数据进行去重处理,可以提升无线传感网络的防御性能。结果如图 8 所示。

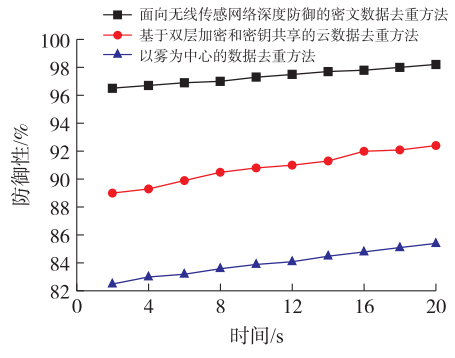


图 8 防御性能对比

分析图 8 可知,随着无线传感网络中密文数据去重时间的增长,三种方法的网络防御性能均有不同程度的提升。其中面向无线传感网络深度防御的密文数据去重方法的防御性一直保持在 96%以上,表明该方法通过密文数据去重可以有效地提高无线传感网络的防御能力。

4　结束语

为了降低重复密文数据对网络生命周期的影响,提出了面向无线传感网络深度防御的密文数据去重方法。仿真结果表明,该方法能有效提高无线传感网络的剩余能量、吞吐量以及空间压缩率,对密文数据的去重效果较好。

参考文献:

[1] Achar S J,Baishya C,Kaabar M K A.Dynamics of the Worm Transmission in Wireless Sensor Network in the Framework of Fractional Derivatives[J]. Mathematical Methods in the Applied Sciences,2022,45(8):4278-4294.

[2] 哈冠雄,贾巧雯,陈杭,等. 无第三方服务器的基于数据流行度的加密去重方案[J]. 通信学报,2022,43(8):17-29.

[3] 高文静,咸鹤群,程润辉. 基于双层加密和密钥共享的云数据去重方法[J]. 计算机学报,2021,44(11):2203-2215.

[4] Yoosuf M S,Anitha R. Low Latency Fog-Centric Deduplication Approach to Reduce IoT Healthcare Data Redundancy[J]. Wireless Personal Communications,2022,126(1):421-443.

[5] 朱海洋,钱中昊,严凡,等. 支持多维度数据去重的交互式可视分析方法[J]. 计算机辅助设计与图形学学报,2022,34(6):841-851.

[6] 潘琢金,蒋欣格,王传云,等. 基于移动代理的 WSN 压缩数据采集方法研究[J]. 计算机仿真,2021,38(4):315-320.

[7] 李梓正,曹司磊,王瑶. 基于协方差矩阵优化重构的稳健波束形成算法[J]. 兵器装备工程学报,2023,44(9):306-312.

[8] Hadri Z E,Capecchi V. Computation of the Covariance Matrix Implied by a Recursive Structural Equation Model with Latent Variables[J]. Quality & Quantity: International Journal of Methodology,2022,56(6):4295-4311.

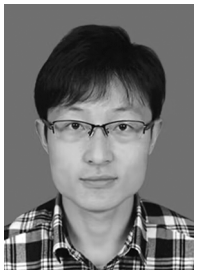
[9] 杨鹏,赵子珩,王中冠,等. 基于状态空间线性变换的主动配电网分布式电压控制[J]. 电力自动化设备,2023,43(1):64-72.

[10] Karuppaiah D,Durai R V P M. Hybrid Approach for Semantic Similarity Calculation Between Tamil Words[J]. International Journal of Innovative Computing and Applications,2021,12(1):13-23.

[11] 赵莉,张克旺. 基于傅里叶描述子的实装战技图像典型特征相似度检测方法[J]. 南京理工大学学报,2022,46(4):406-411.

[12] 王路曼,王理民,蔡铭. 基于多源速度数据的动态权值融合方法研究[J]. 中山大学学报(自然科学版)(中英文),2022,61(5):31-40.

高俊杰(1984—),男,汉族,山西浑源人,现为山西大学实验师,主要研究方向为智能信息处理;



杨　帆(1987—),男,汉族,山西新绛人,现为山西医科大学计算机教学部实验师,主要研究方向为数据挖掘、信息智能处理。