

ترکیب و تجمیع داده در شبکه های حسگر بیسیم Data Fusion/Aggregation

محمد حسین یغمایی مقدم
دانشگاه فردوسی مشهد

مقدمه

- شبکه-های سنسوری بی-سیم (WSN) شامل هزاران سنسور با هزینه و توان پردازشی و مصرفی پایین می باشد.
- به دلیل توان پایین گره-ها توسعه این شبکه ها برای طولانی مدت با مشکل همراه میباشد
- با افزایش کاربردهای شبکه-های سنسوری بی-سیم بایستی بدنبال الگوریتم-ها و پروتکل-هایی بود تا طول عمر شبکه را افزایش دهند.
- **استفاده از پروتکل-های مسیریابی با حداقل مصرف توان**
- **فشرده-سازی داده-ها قبل از ارسال**
- **پردازش سیگنال با همکاری گره-ها**
- **ترکیب داده ها**

مقدمه

- در کاربردهای WSN، تعداد زیادی سنسور در محیط پراکنده شده و پس از وقوع یک رویداد اطلاعات مربوط به رویداد را به پایگاه اصلی ارسال می-کنند.
- پس از وقوع رویداد حجم عظیمی از پیام-ها در شبکه به سمت پایگاه اصلی (BS) جاری می-شود
- اکثر این پیام-ها حاوی اطلاعات تکراری از وقوع رویداد هستند
- این افزونگی در اطلاعات منابع شبکه از جمله پهنای باند و توان را مصرف کرده و بنابراین کارایی و طول عمر شبکه کاهش می-یابد.
- تاثیر مهم این امر در مورد گره-های نزدیک پایگاه اصلی است
- با ارسال پیام های تکراری گره ها، تعداد بیشتری پیام از گره های نزدیک پایگاه اصلی عبور می کنند
- این یعنی اتمام توان گره های نزدیک به پایگاه اصلی، و با وجود اینکه سایر گره-های شبکه هنوز دارای انرژی هستند شبکه از کار می-افتد.

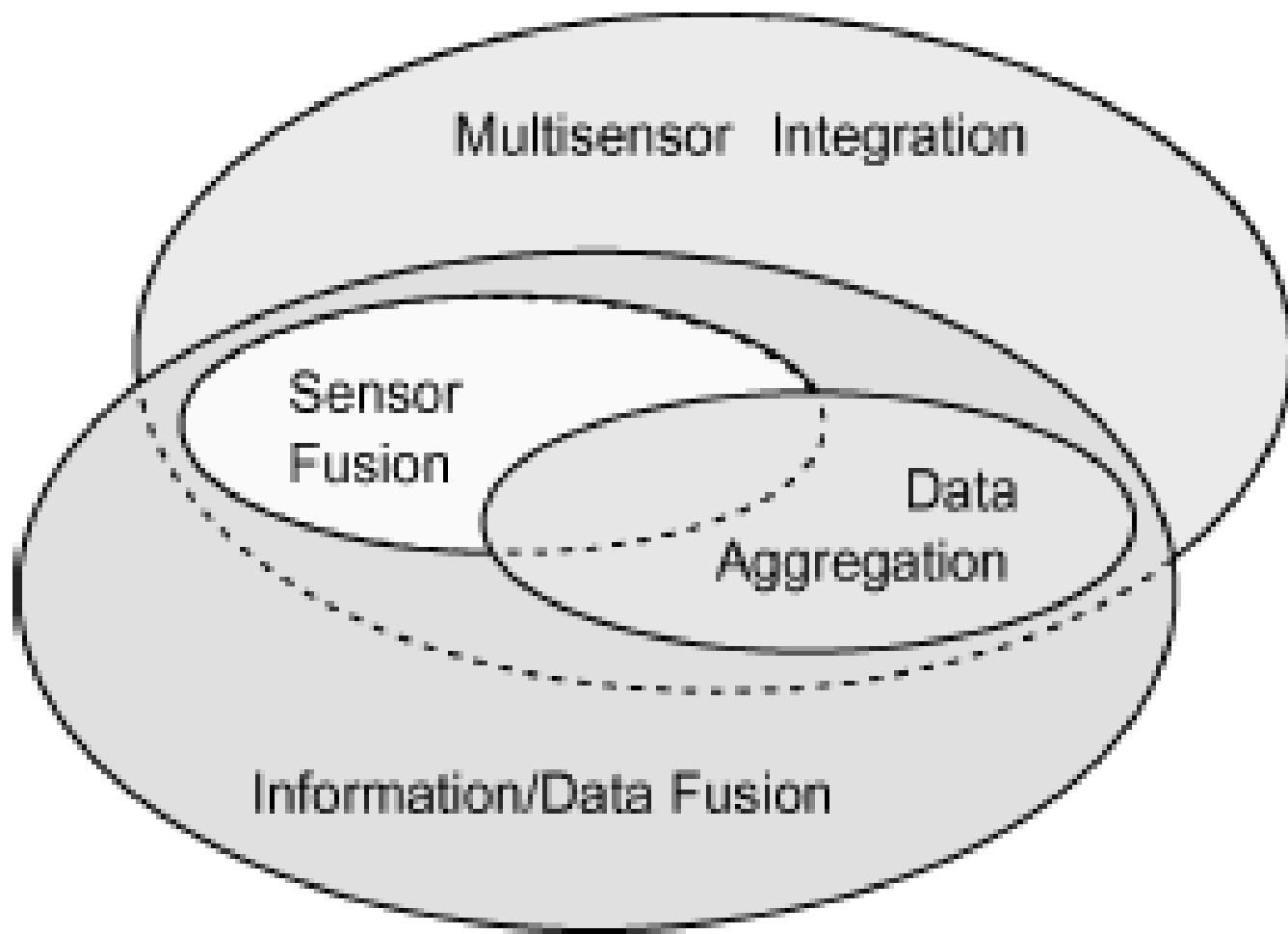
تجمیع داده

- به منظور کاهش تعداد این بسته-ها-گره-ها با همکاری محلی با یکدیگر اطلاعات خوانده شده از محیط را که همراه با افزونگی است، قبل از ارسال پردازش کرده و نتیجه-ی این پردازش را ارسال کنند.
- بدین ترتیب تعداد پیام-های ارسالی کاهش می-یابد.
- این پردازش شامل فشرده سازی، تجمیع و ترکیب اطلاعات می-باشد.

تجمیع و ترکیب داده ها

- تجمیع داده یعنی ترکیب اطلاعاتی که یک نوع دارند ولیکن از منابع متفاوت جمع آوری شده-اند
- چنانچه این داده-ها از یک نوع نباشند، ترکیب نامیده می-شود
- یک گره داده هایی که از چندین گره جمع آوری کرده است ترکیب می کند و نتیجه ی این ترکیب را به پایگاه اصلی ارسال می کند
- بدین ترتیب با ترافیکی شبکه کاهش و در مصرف انرژی صرفه جویی می شود.

تجميع و تركيب داده ها



کاربردهای ترکیب داده ها

• کاربردهای نظامی:

- تشخیص هدف،
- سلاح-های هوشمند،
- وسایل نقلیه-ی خودکار،
- نظارت میدان نبرد
- سیستم های تشخیص تهدید خودکار

• کاربردهای غیر نظامی

- نگهداری ماشین آلات پیچیده،
- رباتیک
- کاربردهای طبی

ترکیب اطلاعات

- ترکیب چندین منبع اطلاعاتی برای بدست آوردن اطلاعات بهبود یافته
- در حوزه ی WSN مکانیزم های تجمیع برای کاهش ترافیک و در نتیجه کاهش مصرف توان شبکه پیشنهاد شده است
- ترکیب داده در تقریب مکان سنسورها، تشخیص شکست در مسیریابی و جمع آوری اطلاعات لینک برای مسیر یابی کاربرد دارد.
- روش های پایه:
min, max, ave

تجمیع داده

- مجموعه ای از روش ها برای ترکیب اطلاعاتی که از چند گره حسگر جمع آوری شده به منظور استخراج اطلاعات با معنا.
- پروسه ای چند سطحی و چند بعدی که در ارتباط است با ترکیب داده ها و اطلاعات از چندین مبدا.
- ترکیب داده از چندین گره حسگر به منظور بهبود دقت و بدست آوردن استنتاج های بهتر نسبت به شرایطی که داده تنها از یک مبدا فراهم آید.

تجمیع داده

- تجمیع داده در WSN بعنوان مترادفی برای ترکیب اطلاعات مطرح شد.
- تجمیع داده شامل:
 - جمع آوری داده های خام از منابع مختلف
 - ترکیب این داده های خام به داده ی تصحیح شده و کم حجم تر
 - تحویل به موقع این داده های به کاربر می باشد
 - ترکیب بایستی انعطاف پذیر و قابل برنامه نویسی باشد

چرا ترکیب اطلاعات

- سنسور ها با محیط در ارتباط هستند و تغییرات در محیط را اندازه می گیرند؛

- **مشکلات:**

- امکان دارد دچار خطا در اندازه گیری شده و یا از کار بیفتد.
- با از کار افتادن هر سنسور ممکن است محدوده ی تحت کنترل شبکه و یا توانایی ارتباط مختل شود.
- محدوده تحت کنترل بایستی به طور کامل پوشش داده و هر تغییری در ناحیه بایستی توسط سنسور ها حس شود

چرا ترکیب اطلاعات

- برای غلبه بر شکست سنسورها، محدودیت های تکنیکی و مسائل مربوط به پوشش، سه ویژگی بایستی در شبکه برآورده شود:
 - همکاری سنسورها
 - افزودنی
 - مکمل بودن

چرا ترکیب اطلاعات

- می توان با قراردادن چندین سنسور به جای یک سنسور ناحیه مورد نظر را پوشش داد.
- گره ها با همکاری با یکدیگر تغییرات ناحیه را حس کرده و بعد با ترکیب این اطلاعات می توان نسبت به ناحیه دید کاملی داشت.
- افزودن WSN در این شبکه نسبت به شکست یک گره آسیب پذیر باشد جلوگیری می کند
- با دریافت ویژگی های متفاوت محیط توسط سنسور ها، می توان این اطلاعات مکمل را با هم ترکیب کرد و نتایج بهتری نسبت به شرایطی که همه داده ها از یک ویژگی رویداد هستند، گرفت.

محدودیت های ترکیب اطلاعات

- ترکیب اطلاعات بایستی به صورت توزیع شده انجام شود
- با در نظر گرفتن بار ترافیکی، سیستم ترکیب متمرکز ممکن است نسبت به توزیع شده کارا تر عمل کند.
- به خاطر محدودیت منابع در WSN الگوریتم های محلی و توزیع شده ارجح هستند
- گمشدن بسته ها در این ارتباطات زیاد است
- علاوه بر این موارد بایستی هزینه ی عمل ترکیب را در نظر گرفت

روش های ترکیب داده ها

- روش های پایه

- average و min, max

- روش های پیچیده:

- روش های Bayesian ،

- Dempster shafer ،

- منطق فازی

- الگوریتم های ژنتیک

دسته بندی ترکیب اطلاعات

- دسته بندی براساس رابطه ی بین منابع

- **مکمل:**

- در این روش، منابع داده هایی مکمل فراهم می آورند و این داده ها مربوط به قسمت هایی مختلف از محیط هستند بعنوان مثال دمای مربوط به شرق و غرب یک ناحیه

- **افزونه:**

- چنانچه دو و یا بیشتر سنسور های مستقل یک اطلاع را بدست آورند این اطلاعات به منظور افزایش اطمینان و دقت با هم ترکیب می شوند و البته حجم ترافیک نیز کاهش می یابد.

- **همکار:**

- دو منبع مستقل داده هایی را فراهم می آورند که اطلاعات جدید و احتمالا پیچیده تری از این داده ها استخراج می شود؛ بعنوان مثال ترکیب داده های فاصله و زاویه برای محاسبه ی مکان شی.

دسته بندی ترکیب اطلاعات

• دسته بندی براساس سطح انتزاع

- ۴ سطح انتزاع برای ترکیب اطلاعات در نظر گرفته می شود:
 - سیگنال، پیکسل، ویژگی و نماد.
- سطح سیگنال با سیگنال های یک و یا چند بعدی سنسور ها کار می کند. این سطح را می توان در کاربرد های بلادرنگ و یا بعنوان مرحله ی میانی در ترکیب های بعدی در نظر گرفت.
- سطح پیکسل در مورد بهبود پردازش تصویر مطرح است.
- در سطح ویژگی با ویژگی هایی از قبیل سرعت و شکل سیگنال ها و تصاویر کار می شود
- در سطح نماد که سطح تصمیم نیز نامیده می شود اطلاعات نمادی از یک تصمیم است.
- دو سطح ویژگی و نمادی در کاربرد های تشخیص شی بکارگرفته می شوند.

دسته بندی ترکیب اطلاعات

- بنا به سطح انتزاع داده، اندازه، ویژگی و تصمیم ترکیب اطلاعات به ۴ دسته تقسیم می شوند؛

• ترکیب سطح پایین :

- این روش ترکیب اندازه ها و همچنین سیگنال ها نیز نامیده می شود. داده های خام با هم ترکیب شده و داده ی دقیق تر را فراهم می کنند بدین دلیل که نویز کاهش می یابد.

• ترکیب سطح میانه :

- ویژگی های ورودی از جمله شکل، بافت، مکان و ... به منظور انجام وظایف دیگری از جمله تشخیص شی با هم ترکیب می شوند. این روش ترکیب ویژگی /خصوصیت نیز نامیده می شود.

• ترکیب سطح بالا :

- ترکیب تصمیم یا نماد نیز نامیده می شود. با ترکیب تصمیم ها سعی بر اینست که تصمیم با اطمینان بیشتر و یا یک تصمیم سراسری اتخاذ شود.

• ترکیب چند سطحی :

- در این جا داده از سطوح مختلف با هم ترکیب می شوند.

دسته بندی ترکیب اطلاعات

- در بعضی موارد، سطح اندازه را حذف کرده و سه سطح ترکیب در نظر می گیرند:
 1. **ترکیب داده:**
 - داده های خام به منظور بدست آوردن داده دقیق تر با هم ترکیب می شوند و بدین ترتیب هزینه ی انتقال روی داده کاهش می یابد و پهنای باند هدر نمی رود.
 2. **ترکیب ویژگی:**
 - ویژگی هایی که هر گره پس از دریافت داده خام از محیط داده استخراج می کنند به مرکز ترکیب ارسال کرده و مجموعه این ویژگی ها در مرکز ترکیب با هم ترکیب می شوند. این سطح ترکیب از دقت اطلاعات می کاهد.
 3. **ترکیب تصمیم:**
 - هر گره به صورت محلی تصمیم خود را گرفته و مجموعه این تصمیم ها در مرکز ترکیب برای اتخاذ تصمیم نهایی با هم ترکیب می شوند.
 - هر چه سطح ترکیب پیچیده تر می شود مصرف توان و همچنین پهنای باند لازم برای ارسال نتیجه ترکیب افزایش می یابد.

الگوریتم ها، تکنیک ها و روش ها

- روش های ترکیب داده را می توان بنا به جنبه های مختلف دسته بندی کرد از جمله:
 - سطح انتزاع داده، هدف، پارامترها، نوع داده و مبنای ریاضی.
- با در نظر گرفتن هدف ترکیب اطلاعات، می توان روش ها را چنین دسته بندی کرد؛
 - استنتاج،
 - تخمین،
 - دسته بندی،
 - نگاشت ویژگی،
 - تجمیع
 - فشرده سازی

استنتاج

- این روش ها در ترکیب تصمیم بکار گرفته می شود.
- در این نمونه بنا به وضعیت های دریافت شده تصمیم اتخاذ می شود.
- انتقال از پیشنهاد درستی به پیشنهاد درست دیگری استنتاج تلقی می شود.
- روش های استنتاج کلاسیک بر مبنای استنتاج Bayesian و تئوری DS است.
- از دیگر روش های استنتاج می توان منطق فازی، شبکه های عصبی، استدلال Abductive و ترکیب معنایی اطلاعات است-Dempster Shafer Belief Accumulation

تخمین

- این روش ها تئوری کنترل را مبنا قرار داده و از احتمالات برای محاسبه اندازه بنا به یک سری اندازه کمک می گیرد.
- روش های تخمین عبارتند از: ML ، MAP ، حداقل مربعات، فیلتر میانگین متغیر، فیلتر کالمن و فیلتر جزئی است.
- Maximum Likelihood
- Maximum A Posteriori
- Moving Average Filter

نگاشت ویژگی

- در بعضی کاربردها از جمله هدایت و مدیریت منابع کار کردن با داده های خام غیر ممکن است.
- در این حالت بایستی از داده های سنسور ها ویژگی هایی را استخراج کرد.
- روش های مختلفی از استنتاج و تخمین را می توان برای نگاشت ویژگی بکار برد.
- از جمله روش های نگاشت می توان Occupancy grid و اسکن شبکه را نام برد.

تجمیع و فشرده سازی

• تجمیع

- تکنیک های تجمیع برای غلبه بر مشکل ازدحام در شبکه و مشترک بودن پیام ها ارائه شدند.
- به دلیل روش های مسیر یابی و همچنین افزونگی سنسورها این مشکلات بوجود می آیند.
- روش های خلاصه سازی، بسته بندی و فرونشانی از این دسته می باشند.

• فشرده سازی

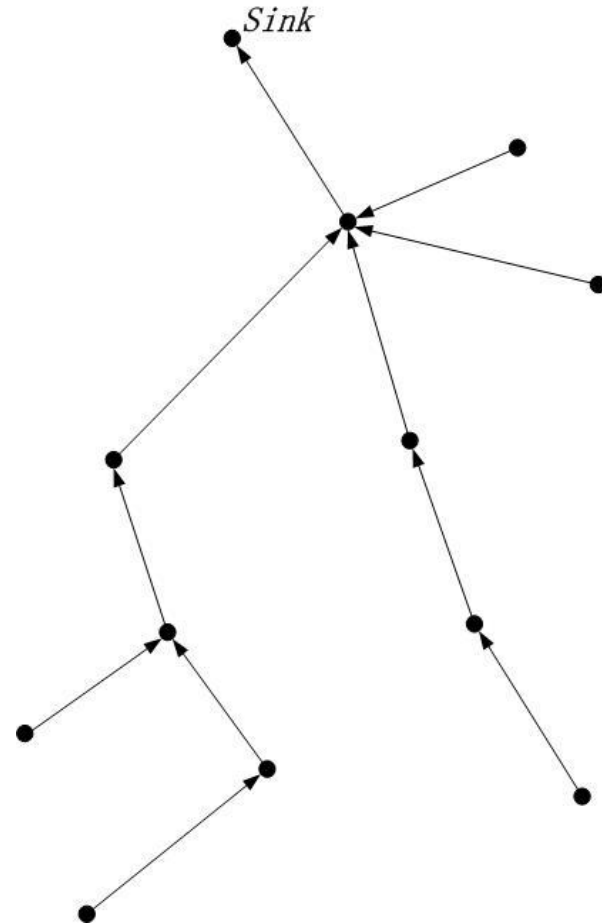
- در WSN داده ها بنا به ارتباطات و همسایگی سنسورها، فشرده می شوند.

عوامل موثر بر تجمیع داده

- هم زمانی گره ها و تاخیر در تجمیع
- زمان بندی در تجمیع داده
- ارسال ترافیک بلادرنگ
- امنیت در تجمیع داده

Data Aggregation Spanning Tree

- Finding the optimal aggregation tree
 - Modeled as a minimum Steiner tree problem
 - An NP-hard problem



Pervious Works

- Sub-optimal solutions[2002]
 - Center at the Nearest Source (CNS)
 - Shortest Path Trees (SPT)
 - Greedy Incremental Trees (GIT)

- Clustering algorithms [2004]
- DCTC [2004]
- Espan [2005]
- LPT [2005]

Pervious Works

- Sub-optimal solutions[2002]
- Clustering algorithms [2004]
 - After partition network into clusters
 - Cluster's members construct aggregation tree and transmit data to cluster head
- DCTC [2004]
- Espan [2005]
- LPT [2005]

Pervious Works

- Sub-optimal solutions[2002]
- Clustering algorithms [2004]
- DCTC [2004]
 - Construct dynamically an aggregation tree
 - Mobile target tracking
 - Subset of nodes participate in tree
- Espan [2005]
- LPT [2005]

Pervious Works

- Sub-optimal solutions[2002]
- Clustering algorithms [2004]
- DCTC [2004]
- Espan [2005]
 - Construct tree for data aggregation
 - By considering node's distance to sink and energy
- LPT [2005]

Pervious Works

- Sub-optimal solutions[2002]
- Clustering algorithms [2004]
- DCTC [2004]
- Espan [2005]
- LPT [2005]
 - Construct tree for data aggregation
 - By considering node's energy

Pervious Works

- Sub-optimal solutions[2002]
- Clustering algorithms [2004]
- DCTC [2004]
- Espan [2005]
- LPT [2005]

پروتکل E-SPAN

- ◆ یک پروتکل توزیع شده است و به منابع در یک حوزه رویداد برای انجام تجمیع داده کمک می کند.
- ◆ در این پروتکل، نود منبعی که بیشترین انرژی باقیمانده را دارد به عنوان ریشه انتخاب می شود.
- ◆ نودهای منبع دیگر، نود والد متناظرشان را از میان همسایه هایشان بر اساس انرژی باقیمانده و فاصله تا ریشه انتخاب می کنند.

پروتکل E-SPAN

- ◆ این پروتکل، یک درخت پوشای آگاه از انرژی را به عنوان درخت تجمیع داده می سازد و یک ریشه مشخصی را برای جمع آوری داده انتخاب می کند.
- ◆ در E-Span، نیازی نیست ریشه، نود مشخصی باشد.
- ◆ برای انتخاب ریشه، انرژی باقیمانده نودها را در حوزه رویداد در نظر می گیریم.

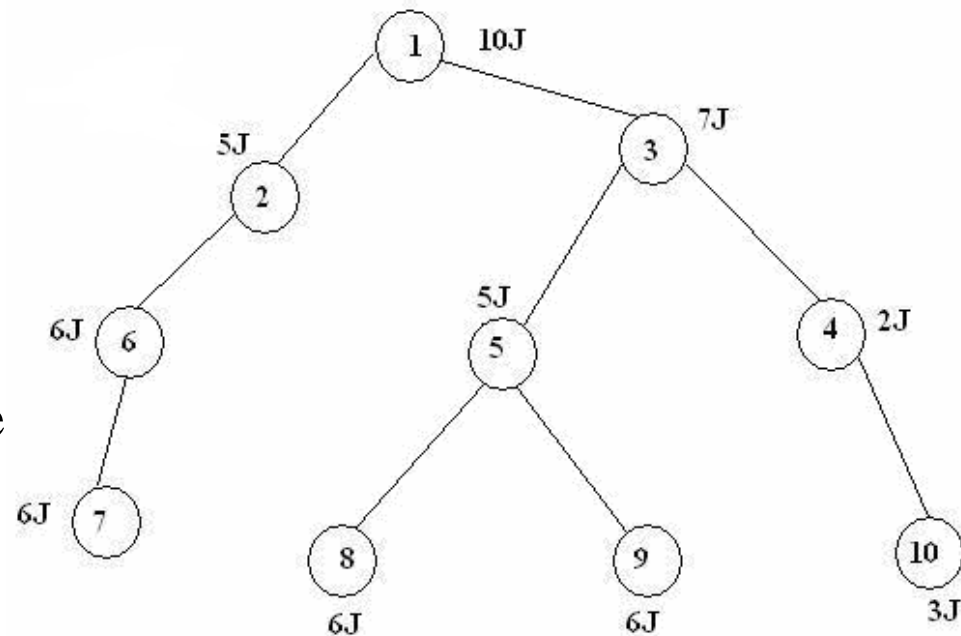
پروتکل E-SPAN

◆ E-Span، متوسط تأخیر انتقال بسته کم و نرخ تحویل بسته بالایی را بدست می دهد.

◆ با استفاده از این پروتکل، عمر منابع به طرز چشمگیری افزایش می یابد.

Espan

- In Espan
 - Each node selects closest neighbors to Sink as parent
- Problem of this algorithm
 - Node with less energy Fails quickly
 - Network can not coverage region completely



LPT

- In LPT
 - Each node selects neighbors with most energy as parent
- Problem of this algorithm
 - Parents may have higher distance to root
 - Cause more energy consumption

