

Software-Infrastrukturen für das Internet der Dinge

Fabian Schlup

ETH Zürich, Mai 2008

Übersicht. Mit der zunehmenden Verbreitung von mit dem Internet verbundenen Sensoren hat man bald die Daten einer grossen Anzahl unterschiedlichster Geräte an verschiedenen Orten der Welt zur Verfügung. Die Aggregation und Analyse von vielen Sensoren in einer gewissen Region und / oder Zeit kann sehr interessante Erkenntnisse bringen, doch bringt die globale Vielfalt von Sensortypen und die riesigen Datenmengen grosse Herausforderungen mit sich. In dieser Arbeit werden drei Software-Projekte vorgestellt, die sich dieser Herausforderungen annehmen und verschiedene Lösungen zur Veröffentlichung, Speicherung, Suche und Verarbeitung von Sensordaten implementiert haben.

1 Einführung

Seit den Anfängen der Computertechnik mit riesigen, zentralen Rechnern ist viel passiert. Heute besitzt fast jeder mindestens einen persönlichen Computer und der Trend zur Einbettung und Integration von Computertechnik in Alltagsgegenständen und der Umgebung zeichnet sich klar ab. Damit sogenannte smarte Dinge den Menschen bei Alltagssituationen bestmöglich und ohne explizite Interaktion unterstützen können, werden oft Sensoren eingesetzt, mit deren Hilfe die Dinge ihre Umgebung selbständig wahrnehmen können. Mit der zunehmenden Vernetzung – getrieben durch die Verbreitung der drahtlosen Kommunikation – bildet sich ein globaler Verbund von smarten Dingen und Sensoren. Dieses „Internet der Dinge“ eröffnet erstmals die Möglichkeit, zahllose Sensormessungen verschiedenster Grössen von unterschiedlichsten Orten für jeden immer und überall verfügbar zu machen. Der fast uneingeschränkte Zugriff auf eine Vielzahl von über ein Gebiet verteilten Sensordaten bietet ungeahnte Möglichkeiten. Unzählige Anwendungsgebiete, von Verkehrsüberwachung, Stauprognosen bis zur Erstellung von Temperaturkarten in Echtzeit (Abbildung 1) sind denkbar. Das Auffinden relevanter Sensoren und der Umgang mit den riesigen und sehr unterschiedlichen Datenmengen, die ein Sensor produziert, stellen allerdings grosse Herausforderungen dar. Zusätzlich stellt sich die Frage, welche Daten man überhaupt verfügbar machen will oder darf, und wie man Zugriffe einschränken kann. Die Entwicklung einer geeigneten Infrastruktur zur Bewältigung dieser Probleme ist ein aktuelles Forschungsthema und Gegenstand mehrerer Projekte.

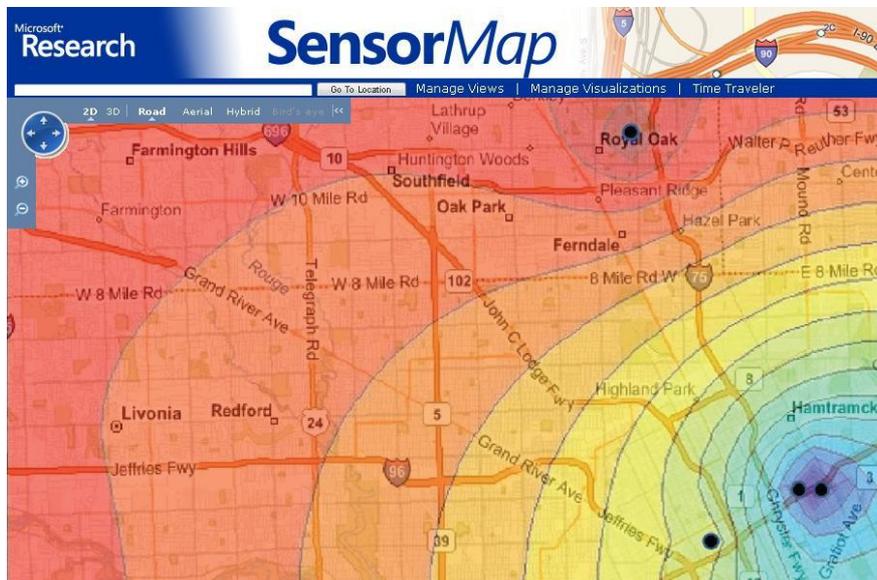


Abbildung 1. Die Applikation SensorMap, basierend auf der Plattform SenseWeb, visualisiert Daten von verschiedenen Temperatursensoren auf einer Landkarte von Windows Live Local

In dieser Arbeit werden drei Software-Projekte vorgestellt, welche die Organisation von Sensoren und die Verarbeitung von Sensordaten zum Ziel haben. Anschliessend werden auf die Herausforderungen im Bereich Sensor-Infrastruktur im Allgemeinen und im Kontext der beschriebenen Projekten im Speziellen eingegangen. Dabei werden die vorgestellten Systeme verglichen und Unterschiede bezüglich Anwendungsgebiet und Lösungsansatz aufgezeigt. Eine kurze Zusammenfassung und Schlussfolgerung bilden den Abschluss.

2 Projekte

2.1 GSN

Die Global Sensor Networks (GSN) Middleware wird an der Ecole Polytechnique Fédérale de Lausanne (EPFL) entwickelt. Die Java-basierende Komponente bietet die Möglichkeit, unterschiedliche Sensoren und Sensornetze einfach und einheitlich ans Internet anzuschliessen, Sensordaten zwischenspeichern, weiterzuverarbeiten und mit anderen Middleware-Instanzen im Netz auszutauschen. Mit ihrer Unterstützung für direkt und indirekt an der Hardware angeschlossene Sensoren, lokale Datenverarbeitung und der abstrahierten Schnittstelle gegen aussen, funktioniert die Software nicht nur als Bindeglied zwischen Sensoren und Internet (Abbildung 2), sondern im Verbund auch als verteilte Berechnungsplattform.

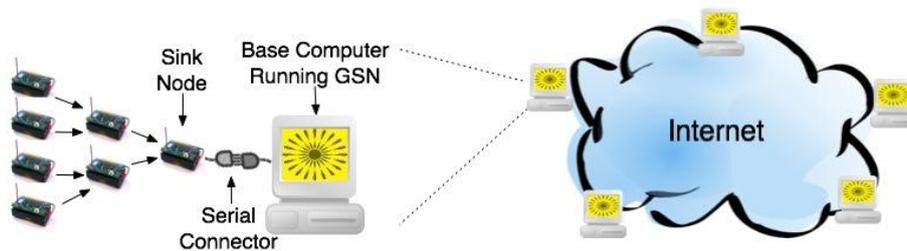


Abbildung. 2. Die GSN-Instanz als Bindeglied zwischen Sensornetz und Internet [7]

Die Grundlage der GSN-Middleware bildet das Konzept der „virtuellen Sensoren“. Ein solcher Sensor ist ein abstraktes Objekt mit einheitlicher Schnittstelle, welches einen Datenstrom erzeugt. Ein Datenstrom wird in GSN als Sequenz von Datentupeln (mit Zeitstempel) abgebildet. Ein virtueller Sensor kann einen direkt am Rechner angeschlossenen, echten Sensor repräsentieren, aber auch eine künstliche, aus einer Aggregation verschiedener Sensordaten entstandene Datenquelle darstellen. Die Definition eines neuen virtuellen Sensors erfolgt ohne Programmieraufwand in Form einer XML-Datei. Dabei wird die Form des generierten Datenstroms als Tupel deklariert, die eingehenden Datenströme aus verschiedenen Quellen spezifiziert und schliesslich die Ausgabe als Aggregation der Eingaben definiert. Zusätzlich können Metainformationen, wie der Standort des Sensors, angegeben werden. Diese Information erlaubt später die einfache Gruppierung und Selektierung der Sensoren nach gewissen Kriterien (z.B. dem Standort). Der Art der Datenströme, welche als Eingabe verwendet werden können, sind dabei kaum Grenzen gesetzt. Standardmässig gibt es Adapter sowohl für den Datenempfang von direkt an der Hardware angeschlossenen Sensoren via Serial- oder USB-Port, als auch für den Datenbezug von netzfähigen Sensoren mittels HTTP-GET und via TinyOS¹-Knoten. Virtuelle Sensoren auf anderen GSN-Instanzen werden mit XML-RPC angesprochen. Es besteht auch die Möglichkeit eigene Adapter in Java zu schreiben. Die empfangenen Daten werden vom System zusammen mit einem Zeitstempel in einer lokalen, relationalen Datenbank abgelegt, wo sie dann von abhängigen virtuellen Sensoren ausgelesen, aggregiert und weiterverwendet werden können. Die Aggregation kann in SQL definiert werden, für komplexere Anwendungen (z.B. zur Erstellung von Graphiken) können auch hier Java-Klassen als „Ausgabeprozessoren“ definiert werden.

Die aktuelle Implementierung der GSN-Middleware (Version 0.95, Februar 2007) kann von Sourceforge² bezogen werden. Sie bietet neben einer Reihe von Standardadaptern für gängige Sensoren eine Administrationskonsole (HTML) und wird mit einer HSQL-Datenbank³ ausgeliefert.

¹ Linux-Derivat, oft als Betriebssystem der einzelnen Knoten in Sensornetzen benutzt

² <http://gsn.sourceforge.net/>

³ Java-basierte, leichtgewichtige Datenbank (<http://www.hsqldb.org>)

2.2 SensorBase.org

Die Plattform SensorBase.org ist eine zentrale Datenverwaltungskomponente, welche einen einfachen und standardisierten Austausch von Sensordaten ermöglichen soll. Dabei haben die Forscher den Begriff „slog“ (Komposition aus „Sensor“ und „Log“) kreiert, um mit dieser verbalen Anlehnung an den bekannten „blog“ zu verdeutlichen, dass sie mit ihrer Plattform die Veröffentlichung von Sensordaten genau so einfach und selbstverständlich machen wollen, wie das Publizieren eines Textes auf dem eigenen Blog. Das Projekt wird an der University of California Los Angeles UCLA entwickelt.

Entstanden ist die auf PHP und MySQL basierte Plattform aus dem Bedürfnis, die von den eigenen, experimentellen Sensornetzen an der Universität gesammelten Daten effizient zu verwalten und Auswertungen zu erleichtern. Früher wurden die von den Sensornetzen produzierten, riesigen Daten-Logs direkt als Textdateien auf diversen File-Servern abgelegt. Bevor Auswertungen möglich waren, mussten zuerst die gewünschten Dateien von verschiedenen Orten zusammengesucht und zusammengeführt, die benötigten Einträge herausgefiltert, die Logs in ein nützliches Format konvertiert und schliesslich in die Auswertungssoftware importiert werden. Die neue Plattform ermöglicht nun die zentrale und strukturierte Haltung dieser Daten und unterstützt die Datensuche nach Standort oder Typ des Sensors und der Zeit.

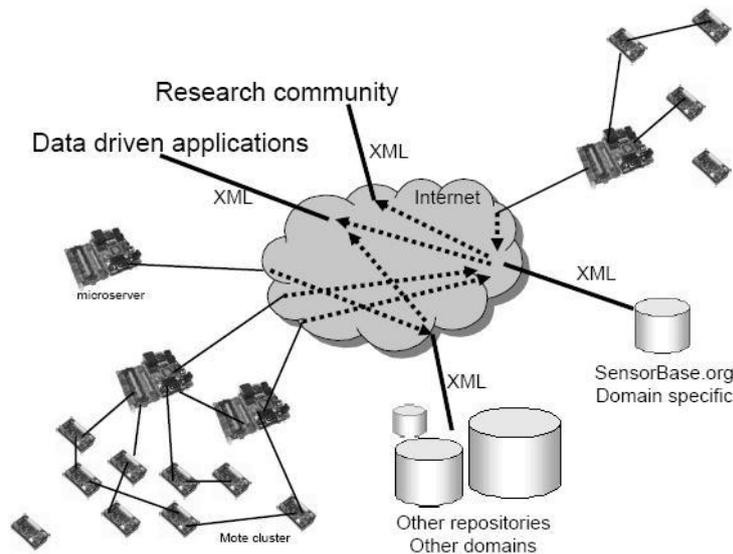


Abbildung 3. Die SensorBase.org-Plattform als zentraler Datenspeicher im globalen Kontext, der Daten von Sensoren und Sensornetzen speichert und zur Verfügung stellt [5]

Die Plattform bietet ein Web-Interface, über das verschiedene Projekte ihre Daten verwalten und Zugriffsrechte vergeben können. Nachdem das Format der

Datenlieferung vom Benutzer definiert worden ist, können Sensoren und Sensornetze mit Internetverbindung neue Daten direkt per HTTP-POST an den Server übermitteln. Datenabfragen können interaktiv via Web-Interface erfolgen oder mit Hilfe eines angebotenen Web-Services (SOAP). In beiden Fällen können die gewünschten Daten mittels SQL bezüglich Sensorstandort, Sensortyp und Zeitraum eingegrenzt und allenfalls bereits zu einem gewissen Grad aggregiert werden. Die Auswertung erfolgt dann mit einer Drittanwendung nach Wahl.

SensorBase.org wird vom Center for Embedded Networked Sensing (CENS) an der UCLA gehostet und steht registrierten Benutzern (Registrierung auf Anfrage) zur Verfügung.

2.3 SenseWeb

Eines der am weitesten fortgeschrittenen Infrastrukturprojekte für das „Sensor Internet“ ist SenseWeb von Microsoft Research. Der Prototyp umfasst verschiedene Komponenten und der Funktionsumfang reicht von Discovery, Datenverwaltung, Suche, Caching und Aggregation bis zu geografischen Visualisierungen. Das erklärte Ziel ist es, eine zentrale Infrastruktur für Sensor-Applikationen zur Verfügung zu stellen. Diese soll Sensordaten effizient und gemäss den individuellen Bedürfnissen der Applikationen zur Verfügung stellen, indem für jede Anfrage automatisch optimale Datenquellen (Sensoren) ausgewählt, die Daten zusammengetragen und allenfalls bereits aggregiert werden. Mit SensorMap haben die Forscher auch gleich die erste auf der neuen Plattform basierende Applikation in Betrieb genommen (Abbildung 1).

Das Herzstück des Systems bildet der Coordinator (Abbildung 4). Über eine Web-Service-Schnittstelle können hier beliebige Sensoren aus aller Welt registriert und so dem System bekanntgemacht werden. Der Coordinator ist auch die zentrale Anlaufstelle für Drittapplikationen. Seine Aufgabe besteht darin, die ankommenden Anfragen gemäss der verfügbaren Sensoren bestmöglich zu beantworten. Dabei geht man davon aus, dass die meisten Abfragen nicht auf einen konkreten Sensor als Datenquelle zielen, sondern eine komplexere Einschränkung nach Typ, Ort und Zeitfenster aufweisen. Um die benötigten Daten zusammenzutragen steht eine grosse Datenbank (SenseDB) zur Verfügung. Diese funktioniert einerseits als Register der vorhandenen Sensoren, deren Standorte und Typen, andererseits aber auch als riesiger Cache für die Sensordaten selbst. Anfragen mit überlappenden geographischen und temporalen Kriterien können so zu einem wesentlichen Teil direkt aus dem Cache beantwortet werden. Bei Anfragen, die nicht aus dem Cache beantwortet werden können, fordert der Coordinator die benötigten Daten von den zuständigen Sense-Gateways (Abbildung 4) an.

Die verteilte Komponente Sense-Gateway bildet die Schnittstelle zwischen dem zentralen Coordinator und den individuellen, lokalen Infrastrukturen der Sensorbetreiber. Jeder Datenlieferant kann sein eigenes Sense-Gateway mit vorgegebener Web-Service-Schnittstelle für den Coordinator implementieren und lokal betreiben. Dabei ist es dem Datenlieferant überlassen, ob ein Gateway als lokaler Zwischenspeicher realisiert wird, der Daten der lokalen Sensoren sammelt und

aufbereitet, oder ob die Komponente lokale Sensoren auf Anfrage anspricht. Um den relativ grossen Aufwand einer Eigenentwicklung in diesem Bereich für kleine Anbieter zu vermeiden, wurde ein öffentliches Gateway – das DataHub – entwickelt. Einmal registriert, übernimmt das DataHub die Kommunikation mit einem internettauglichen Sensor automatisch. Ist der Sensor in einem privaten Netz von einer Firewall oder einem Router abgeschirmt, so können Daten auch aktiv via Web-Service an das DataHub geliefert werden.

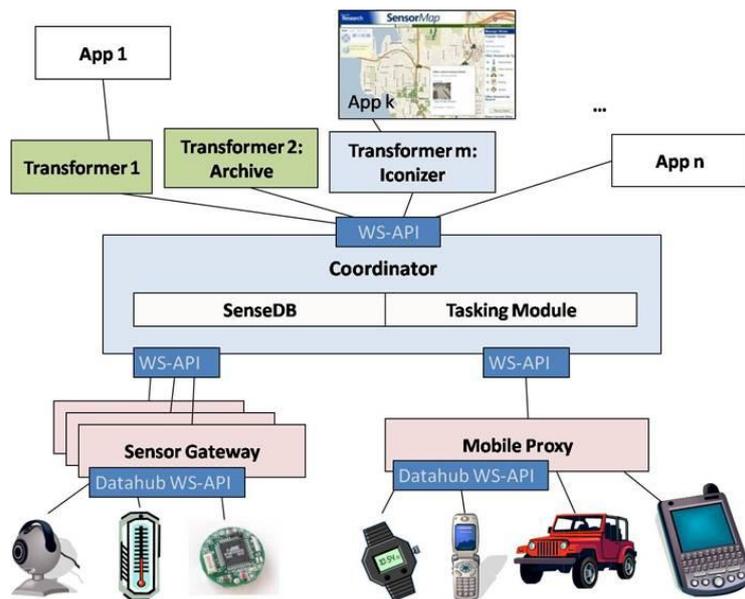


Abbildung. 4. Die Architektur von SenseWeb basiert auf dem zentralen Coordinator, verteilten Sense-Gateways / Mobile Proxys und den Transformers [4]

Die letzte grosse Komponente des SenseWeb-Systems bilden die Transformers. Sie werden benützt um abgefragte Sensordaten zu aggregieren und konvertieren. Die Transformers sind – kategorisiert nach Eingabe- und Ausgabeformat – im Coordinator indexiert und für Drittapplikationen frei benutzbar. So können zum Beispiel Transformers zur graphischen Darstellung verschiedener Datenströme bereitgestellt werden, mit deren Hilfe auch Laien ohne Fachwissen über die zu grundlegenden Datenstrukturen rasch eine Übersicht gewinnen können.

Die beschriebene SenseWeb-Plattform ist zur Zeit im Entstehen und noch nicht öffentlich verfügbar. Das Vorgängerprojekt SensorMap, welches trotz Architekturunterschiede bereits einen Grossteil der beschriebenen Funktionalität abdeckt, wird von Microsoft Research gehostet (Version 3, April 2008) und kann frei benutzt werden⁴. Es wurden bereits diverse Datenlieferanten eingebunden und es

⁴ <http://atom.research.microsoft.com/sensewebv3/sensormap/>

existieren auch bereits Drittprojekte, die auf SensorMap basierende Applikationen entwickeln.

3 Herausforderungen, Probleme und Lösungsansätze

Das Auffinden, Organisieren, Zusammentragen und Verarbeiten einer gewaltigen Menge von sehr unterschiedlichen, verstreuten und häufig auch sensitiven Daten, so wie sie Sensoren generieren, stellt eine grosse Herausforderung dar. Im Folgenden werden einige der wichtigsten Probleme vorgestellt und die Lösungsansätze der besprochenen Projekte verglichen.

3.1 Discovery

Für Suchmaschinenbetreiber besteht eine Kernaufgabe in der automatischen Entdeckung und Registrierung neuer Webseiten. Das Grundprinzip besteht dabei aus dem systematischen Verfolgen von Hyperlinks. Auch im Internet der Dinge wäre eine automatische Auffindung von neuen Sensoren hilfreich, stellt aber angesichts der fehlenden Links eine Herausforderung dar. Zudem möchte man gefundene Sensoren nicht bloss registrieren, so wie das im Web der Fall ist, sondern auch ansprechen können. Letzteres würde bedeuten, dass Sensoren in der Lage sein müssten, sich selbst zu beschreiben.

Die GSN-Middleware macht sich die zunehmende Verbreitung des Standards IEEE 1451 zu Nutze. Dieser Standard definiert das Format eines sogenannten Transducer Electronic Data Sheet (TEDS), welches – eingebaut in Sensoren aller Art – die Messgrösse, Datentyp, Kalibrierung und weitere Informationen über die Komponente beinhaltet. Sobald ein neuer Sensor in die Reichweite eines bei GSN konfigurierten Sensornetzes kommt, reagiert die Middleware mit der automatischen Auslesung dieser Daten und der dynamischen Generierung einer neuen virtuellen Sensor-Instanz.

Einen eher traditionellen Ansatz versuchen die Forscher des SenseWeb-Projekts. Ein Prototyp zur Auffindung von Live-Kameras durchsucht das Internet nach Webseiten mit charakteristischen Merkmalen. Wenn der Algorithmus glaubt, eine eingebundene Live-Übertragung gefunden zu haben, wird die URL registriert und versucht, Metainformationen wie den Standort aus nahegelegenen Textpassagen herauszufiltern. Der Prototyp ist vorerst lediglich für Verkehrsüberwachungskameras ausgelegt und operiert in einem sehr eingeschränkten Bereich (z.B. Webseiten des Verkehrsdepartements von Washington DC).

3.2 Heterogenität

Heute existieren dutzende Arten von elektronischen Sensoren. Es gibt Thermometer, Barometer, Voltmeter, Foto- und Videokameras, Lichtschranken, GPS und Messeinheiten zur Bestimmung von Lichtstärke, Luftverschmutzung oder Herzfrequenz. Dabei messen die verschiedenen Sensoren natürlich sehr

unterschiedliche Grössen, die sowohl in der Form (Skalar, Vektor, Bild, Video) als auch in der Bedeutung (z.B. Normalbild und Wärmebild) erheblich voneinander abweichen können. Des Weiteren unterscheiden sich die Geräte in Bezug auf Genauigkeit, Ausgabeformat und andere Details von einem Hersteller zum nächsten. Einige sind netzfähig und gewähren Zugriff via TCP/IP oder senden Daten aktiv weiter, andere müssen per USB direkt an einen Rechner angeschlossen werden. Einige Geräte, wie Überwachungskameras, können sogar aktiv angesteuert werden und bieten so die Möglichkeit den Bereich der Aufzeichnung individuell zu wählen. Um einen optimalen Nutzen für Auswertungen aller Art zu erzielen, müssen die Daten von möglichst vielen und verschiedenen Sensoren im „Sensor-Internet“ kombiniert werden können. Doch damit das Gelingen kann, muss man einen Weg finden, wie man mit der Vielzahl unterschiedlichster Sensoren und den Daten, die sie produzieren, einheitlich umgehen kann.

Die virtuellen Sensoren sind der Ansatz des GSN-Projekts für dieses Problem. Hinter dieser einheitlichen Fassade können Unterschiede bezüglich Kommunikation mit dem Sensor und Format der Rohdaten versteckt werden. Beim Eintreffen neuer Daten vom Sensor (je nach Implementierung entweder Empfang eines „Push“ vom Sensor oder ein aktiver „Pull“ des GSN-Adapters) werden diese in ein vorher definiertes Datentupel umgewandelt und in die Datenbank eingetragen. Danach werden Auswertungen (ebenfalls in Form virtueller Sensoren), die auf diesen Daten beruhen, automatisch angestossen. Auf Grund von fix definierten Meta-Informationen können virtuelle Sensoren aller Art einheitlich gesucht und für Auswertungen ausgewählt werden.

Bei SensorBase.org wird von Haus aus nur ein Kommunikationsweg unterstützt: HTTP-POST. Die Plattform bleibt dabei passiv, der Sensor muss die Daten „pushen“. Das Format der gelieferten Daten kann vorher definiert werden, damit das System die Konvertierung in die einheitliche Repräsentation vornehmen kann, bevor die Daten in der Datenbank abgelegt werden. Zur Selektierung von Sensordaten dienen auch hier vorgängig erfasste Metadaten, wie Sensortyp und Standort.

Bei SenseWeb werden technische Unterschiede bei der Anbindung der Sensoren am Zentralsystem von den Sense-Gateways behandelt. Wie erwähnt, wird die Kommunikation mit individuellen Sensoren von den Gateways übernommen und der Zugriff für den Coordinator einheitlich als Web-Service zur Verfügung gestellt. Beim Format der Daten unterscheidet SenseWeb zwischen Binär (z.B. Bild oder Ton) und Skalartupel (z.B. Temperatur, räumliche Ausrichtung in drei Dimensionen). Mit den Transformers bietet SenseWeb ein Konstrukt, das es erlaubt, unterschiedlichste Daten einheitlich zu aggregieren. Ein Benutzer kann sich so zum Beispiel alle verfügbaren Sensoren in einem bestimmten Gebiet – unter Anwendung der jeweiligen Transformers – als Icons auf einer Karte anzeigen lassen. So werden ihm die aktuellen Sensorwerte grafisch, für jede Art von Sensor individuell und sinnvoll, angezeigt, ohne dass vorherige Kenntnisse über die Art von Sensoren nötig waren.

3.3 Sicherheit und Datenschutz

Wo immer Daten im öffentlichen Raum gesammelt werden, kommt unweigerlich die Thematik der Privatsphäre und des Datenschutzes auf. Speziell wenn es um informationsreiche Daten, wie Video- oder Bildaufzeichnungen geht, stellt sich die Frage, ob man die Daten in ihrer Originalform öffentlich zugänglich machen will und darf. Bei firmeninternen oder privaten Sensoren stellt die ungewollte, öffentliche Zugänglichkeit der Daten ein Sicherheitsproblem dar. Deshalb sind Mechanismen zum effektiven Schutz privater Datenströme von zentraler Bedeutung. Anschliessend wird auf die technischen Möglichkeiten der vorgestellten Projekte eingegangen. Dabei ist zu beachten, dass die absichtliche Veröffentlichung von Daten, die gegen die Privatsphäre oder den Datenschutz Dritter verstösst, grundsätzlich in jedem Fall möglich ist. Die damit verbundene Problematik wird hier aber nicht weiter behandelt.

Die GSN-Middleware bietet die Möglichkeit der Zugriffsbeschränkung auf Applikationsebene oder einzelner virtueller Sensoren an. GSN-Instanzen im internen Netz können zusätzlich mit Hilfe einer Firewall gegen Zugriffe von aussen abgeschirmt werden. Alternativ könnte man sensible Daten auch durch Aggregation entschärfen. So könnte man sich vorstellen, ein Überwachungsbild eines Raumes mit einem einfachen Bildverarbeitungsalgorithmus in ein einziges Prädikat „hell“ oder „dunkel“ zu verwandeln. Man könnte dann solche unkritischen Informationen doch noch mit der Welt teilen und das Sensor-Internet bereichern. Im GSN-System liesse sich dies mit einem virtuellen Sensor gut realisieren.

Die SensorBase.org-Plattform stellt keine Möglichkeit zur Verfügung, Daten bereits vor einem Zugriff automatisch zu aggregieren. Dafür können individuelle Zugriffsrechte für jeden Benutzer vergeben werden, womit sensible Daten von Zugriffen durch Dritte geschützt werden können. Der Transport der Daten beim Publizieren durch die Sensoren und der Abruf durch die User (SOAP via HTTP) ist allerdings in der aktuellen Version ungeschützt. Dies könnte aber einfach unter Benützung von TLS/SSL realisiert werden.

Vielfältige Möglichkeiten bietet SenseWeb mit seinem Sense-Gateway-Konzept. Weil der Datenlieferant sein eigenes Sense-Gateway implementieren und betreiben kann, liegt die Kontrolle darüber, welche Daten in welcher Form das interne Netz verlassen, gänzlich bei ihm. So sind sogar Zugriffsbeschränkungen abhängig von der Uhrzeit oder – mittels automatischer Bilderkennung – abhängig von der Präsenz eines Menschen in einem Videobild denkbar. Das öffentliche DataHub – das Sense-Gateway für jedermann – bietet dem Datenlieferant Zugriffskontrolle auf Benutzerebene. Während Daten so zwar vor dem Zugriff Dritter geschützt werden können, haben sie in diesem Szenario aber bereits das Internet (verschlüsselt) durchquert und liegen nun an einem zentralen Ort ausserhalb des eigenen Einflusses.

3.4 Leistung und Skalierbarkeit

Mit der verteilten Architektur der GSN-Middleware können hier die Lasten geteilt und die einzelnen Knoten im Vergleich zu einem zentralen System entlastet werden. GSN verfolgt einen Peer-to-peer-Ansatz, bei dem alle GSN-Instanzen ihre Daten lokal

abspeichern und zur Verfügung stellen, aber auch selbst Daten von anderen konsumieren und aggregieren. Die Middleware dient also nicht nur der Anbindung der lokalen Sensoren ans Netz, sondern auch als verteilte Auswertungsplattform. Eine grosse Rolle bezüglich Zugriffszeiten spielt natürlich die Datenbank. Dabei ist ein wichtiger Punkt, dass die GSN-Instanzen, bei Bereitstellung der passenden JDBC-Treiber, mit einem beliebigen relationalen Datenbanksystem betrieben werden können. Bei Lasttests mit MySQL 5.18 mit bis zu 500 Benutzer hat das System gute Skalierbarkeit bewiesen.

Bei SensorBase.org ergeben sich aus der zentralen Datenhaltung grössere Herausforderungen betreffend Leistung und Skalierbarkeit als bei einem verteilten System. In seiner gegenwärtigen Version kommt die Plattform mit einer einzigen, indextierten Datenbanktabelle für die Datenhaltung aus. Diese ist nicht normalisiert und sieht einen einfachen Wert pro Zeiteinheit und Sensor vor. Ein Lastentest mit bis zu 97 Millionen Datenzeilen hat erhebliche Leistungs- und Skalierbarkeitsprobleme zu Tage gebracht. Eine Massnahme, welche die Forscher für nächste Versionen in Betracht ziehen, sind die Verwendung von separaten Datenbanktabellen pro Projekt.

Die Beteiligten des [SenseWeb-Projekts](#) waren sich der Problematik der enormen Datenmengen von Anfang an bewusst, und haben die performante Datenverwaltung ins Zentrum ihrer Bemühungen gestellt. Den Lastproblemen, wie sie bei SensorBase.org auftreten, wird hier mit zwei Techniken begegnet: Auslagerung der Datenarchive und Caching. SenseWeb betreibt grundsätzlich keine zentrale Datenspeicherung. Der Coordinator wählt die nötigen Sensoren zur Beantwortung einer Anfrage und holt sich die Daten dann über das zuständige Sense-Gateway jedes Sensors. Um die Antwortzeiten dennoch möglichst kurz zu halten, werden die Rückmeldungen der Gateways zentral zwischengespeichert und folgende Anfragen nach Möglichkeit (und unter Berücksichtigung einer Verfallzeit) direkt aus diesem Cache beantwortet. Die Cachezugriffe werden zusätzlich optimiert, indem in geographischen Regionen mit einer hohen Dichte von relevanten Sensoren Daten zusammengefasst werden. Diese aggregierten Resultate für eine bestimmte geographisch und zeitlich eingeschränkte Datenbasis werden ihrerseits ebenfalls gecached und für spätere Abfragen und Zusammenzüge über noch grössere Bereiche wiederverwendet. Die Implementierung dieses Verfahrens mit einem COLR-Baum (eine Variante des R-Baumes) auf Microsoft SQL Server hat bei 400'000 Sensoren auf bis zu 100 Anfragen pro Sekunde skaliert. Die Forscher sehen weiteres Potential in der Verteilung des gegenwärtigen Datenbanksystems auf mehrere Server.

4 Zusammenfassung und Fazit

Die vorgestellten Projekte und ihre Software-Prototypen haben Gemeinsamkeiten aber auch viele Unterschiede. Während GSN im Wesentlichen eine Middleware-Komponente für die Anbindung von lokalen Sensoren mit dem Netz ist, konzentrieren sich SensorBase.org und SenseWeb auf die globale Organisation von Sensordaten. Dabei setzt SensorBase.org auf die zentrale Speicherung aller Daten und die aktive Anlieferung der Daten durch die Erzeuger. SenseWeb basiert dagegen auf einer

verteilten Datenhaltung mit zentralem Cache, der Daten bei Bedarf (lazy) beschafft. Jetzt wollen wir die drei Projekte kurz zusammenfassen und die Alltagstauglichkeit bewerten.

Die GSN-Middleware ist eine praktische Komponente zur Anbindung, Verarbeitung und Publizierung von Sensordaten. Die einfache Installation, Konfiguration ohne Programmieraufwand und gute Skalierfähigkeit machen das Produkt einsatztauglich. Dabei muss man sich bewusst sein, dass man wohl Daten von Drittquellen (Quelle kann fremde GSN-Instanz sein oder etwas anderes) für Aggregationen einbeziehen kann, dafür aber in jedem Fall der genaue Endpunkt (URI) bekannt sein muss. Die Komponente ist also vor allem für den Einsatz in einem schon bekannten Umfeld geeignet, das heisst in erster Linie sinnvoll für Sensorbesitzer zur Anbindung der eigenen Sensoren ans Netz (inkl. Aggregationen).

Die Plattform SensorBase.org erfüllt mit den starken Leistungs- und Speicherlimitierungen und der Sicherheitsproblematik bei der Übermittlung sensibler Daten die hohen Anforderungen an ein zentrales System noch nicht und kommt für den operativen Einsatz im grösseren Rahmen wohl kaum in Frage. Man kann weiter davon ausgehen, dass eine zentrale Datenhaltung von rohen Sensordaten für eine globale Lösung angesichts der enormen Datenmengen beim aktuellen Stand der Technik generell unmöglich ist [3]. Auf Grund dieser Tatsachen muss SensorBase.org als Plattform für klar begrenzte Projekte mit unkritischen Daten gesehen werden, was auch dem vorhandenen Konzept der Zugriffskontrolle gegliedert nach Projekten und Projektmitgliedern entspricht.

SenseWeb ist dank der durchdachten, offenen Architektur, der guten Leistungs- und Skalierungseigenschaften, und der Unterstützung für interaktive Datenbezüge auf Anfrage (bei der zum Beispiel eine Kamera als Resultat einer Abfrage die Ausrichtung anpasst) eine ideale Plattform zur globalen Organisation von Sensordaten. Es gibt die Möglichkeit eigene, bereits vorhandene Systeme mittels Web-Service einzubinden, oder bereitgestellte Standardkomponenten zu benutzen. Die Zugriffskontrolle kann in die eigene Hand genommen werden und die Datenströme zwischen Gateways und Coordinator sind in jedem Fall verschlüsselt. Der Heterogenität der Daten und die daraus entstehenden Herausforderungen für die Datenkonsumenten kann durch die Vordefinition und Bereitstellung von Transformers begegnet werden. Für die Organisation der Datenströme im internen Netz der Datenlieferanten (bis zum eigenen Sense-Gateway) wäre hier auch der zusätzliche Einsatz der GSN-Middleware denkbar. Allerdings können dann auf Grund des Prozessierungsmodells von GSN keine Datenabfragen auf Verlangen unterstützt werden.

Das simple Publizieren, wie auch das einfache Suchen und Abrufen von gewünschten Sensordaten auf globaler Basis sind die zentralen Erfolgsfaktoren für eine reiche Applikationslandschaft auf der Basis von Sensordaten. Die Information, die aus der Aggregation von vielen Sensoren über ein gewisses Gebiet und / oder Zeit gewonnen werden können, sind extrem vielfältig und wertvoll. Und schliesslich hängt die Zukunft von smarten Dingen, die auf Sensoren als ihre „Sinnesorgane“ angewiesen sind, massgeblich davon ab, ob sie bald „Ohren und Nasen“ auf der ganzen Welt zur Verfügung haben werden, und so noch mehr und komplexere Aufgaben ohne Rückfragen erledigen können.

Literatur

1. Karl Aberer, Manfred Hauswirth, Ali Salehi: Infrastructure for data processing in large-scale interconnected sensor networks. *Mobile Data Management (MDM), Germany, 2007.*
2. Unbekannter Autor: The Book of GSN (work in progress). <http://gsn.sourceforge.net>, April 15, 2008.
3. Suman Nath, Jie Liu, and Feng Zhao: Challenges in Building a Portal for Sensors World-Wide. *First Workshop on World-Sensor-Web: Mobile Device Centric Sensory Networks and Applications (WSW'2006), Boulder CO, Oct 31, 2006.*
4. Aman Kansal, Suman Nath, Jie Liu, and Feng Zhao: SenseWeb: An Infrastructure for Shared Sensing. *IEEE Multimedia. Vol. 14, No. 4, pp. 8-13, October-December 2007.*
5. Kevin Chang, Nathan Yau, Mark Hansen, and D Estrin: SensorBase.org - A Centralized Repository to Slog Sensor Network Data. *In: Center for Embedded Network Sensing. Paper 116.*
6. MSR SenseWeb Project Website. <http://research.microsoft.com/nec/senseweb>, April, 2008
7. GSN Website. <http://gsn.sourceforge.net>, April 2008
8. SensorBase Website. <http://www.sensorbase.org>, April 2008