Abstract—This paper presents our preliminary results from a single case study on face-to-face, email and instant messaging (IM) communication in a distributed agile software development (DASD) project. The project had 13 team members during the time of study, and it was distributed between three different sites: Finland, Norway and Czech Republic. The team was using Scrum process model with three week sprints.

We studied face-to-face communication with sociometric badges and email and IM communication by analyzing the project-related communication logs provided by each individual team member. We were interested both in the amount of communication and the communication patterns via different communication media.

We found differences in the daily variation of communication amounts across different media as well as different network structure of communication with different media. Specifically, we found some evidence that IM is used as a replacement for face-to-face discussion in DASD, and that email seems to be more suitable for team-wide, inter-site communication than face-to-face or instant messaging.

I. INTRODUCTION

Communication is one of the main success drivers in software engineering [14]. Physical distance reduces the frequency of face-to-face communication [1]. Infrequency of communication, less communication opportunities as well as lower communication richness of mediated communication may lead to communication problems to be emphasized in distributed software development [9], [13], [17].

As agile software development practices emphasize the need for face-to-face communication, we are interested whether mediated communication tools, such as email and IM, could substitute or augment the lack of face-to-face communication in distributed agile projects.

We studied face-to-face, email and instant messaging (IM) communication in a distributed agile software development project. The research problem for the study is: How much and in which way distributed agile software development projects communicate face-to-face, via email and via instant messaging?

The research questions (RQs) for the study are:
1) How much communication takes place face-to-face, via email and via IM in distributed ASD projects?
2) Do the communication patterns differ between face-to-face, email and IM communication?

II. PREVIOUS WORK

A fairly large share of software developers’ time is spent on communicating with other developers in the team. Perry, Staudenmayer and Votta report in their study developers spending on average 75 minutes each day in “unplanned interpersonal interaction” [20]. Frequency of communication among engineers drops radically based on distance, and the frequency being almost at the same low level whether engineers are located 30 meters or miles apart [1]. Earlier studies have reported the slower pace of GSD projects mainly to be caused by challenges and delays due to interaction over distance [7], [9], [10]. A survey on global software development companies studied the problems and challenges in distributed projects. The study classified problems into eight categories. It was found that 74% of the problems in distributed projects were caused by “communication and contacts” [13].

Distributed software development projects need to use different communication tools to overcome the lack of direct face-to-face communication between sites. These tools include e-mail, IM, telephones, teleconferencing, video conferencing, groupware applications, version control systems, integrated development environments and shared workspaces [3], [4], [8], [21], [22], [15].

Agile software development practices emphasize the need for communication both within the software development team and between the team and project’s customers [2]. Agile software development practices claim to solve many issues of software engineering, including long development times, higher costs than anticipated and unmet requirements [11]. While agile software development practices emphasize the need for face-to-face communication, agile projects often also have to use electronic communication tools, as especially in distributed projects not all team members or stakeholders can be present at the same location at once. Empirical studies on existing agile GSD projects suggest that agile methods can be used on GSD projects [18], [19], and can help avoid some of the most common problems of distributed software development [19].

Previous work has been made in studying communication in various contexts of knowledge work. Earlier studies have reported relation between outcome factors, such as performance
and creativity, and communication behavior [6], relationship between face-to-face communication and productivity in IT configuration tasks [24], and the differences in media use and communication patterns across face-to-face, email and IM [12]. However, the presented research contributes on the study of the relationship between face-to-face and electronic communication in distributed agile software development contexts.

III. METHODOLOGY

Following sections describe the case project, and how data was collected and analyzed in this study.

A. Case project

The case project was a distributed software development team of 13 team members at the time of the study. It was distributed across three sites, with sites in Finland, Norway and Czech Republic. The goal of the project was to develop an information management system. At the time of the study, the product was mainly aimed to fulfill in-house demands, and was planned to be taken into use internally. However, there were plans to productize the software for external customers as well.

The Finnish site was mainly responsible for product management: product owner, proxy customer and one developer were located in Finland. Project management and 6 developers were located in Norway, and 4 developers in Czech Republic. All developers at three sites formed a single Scrum team. During the study, we were able to observe two successive three-week sprints. For the first sprint, one team member from Czech Republic site was acting as Scrum master for the team, while the official Scrum master from Norwegian site was on leave.

The team was having daily Scrum meetings every day via teleconference and video conference. In addition to daily meetings, the team had sprint demonstration and planning meetings at the end of each sprint. These meetings were held over video conference, with desktop sharing.

B. Data collection

1) Email and IM communication: In order to get a comprehensive view on the communication networks and patterns of the team, we collected the email archives and chat logs of the team. Electronic communication data was collected by gathering all email and IM communication between team members. The project team used Microsoft Outlook for email and Microsoft Office Communicator for IM. As we were not allowed to install any additional software to corporate servers or individual workstations, we asked all team members to store all project-related emails and instant messages in a specific folder during the study.

As participants voluntarily chose which messages and discussions to store on that folder, they were also able to control themselves which conversations they want to disclose with researchers. While we are not aware any such censorship of the data from the participant side — in fact, most participants just gave us full access to their inbox and outbox — we still want to highlight this, as it is also both potential source of bias as well as an important ethical aspect for other researchers planning and executing similar studies.

Collecting email and IM communication allowed us to objectively evaluate the extent in which email and IM was used in the project studied, to identify actual communication behavior and patterns in mediated communication, and the differences in email and IM use between project sites.

2) Face-to-face communication: We used sociometric badges developed at the MIT Media Lab [16], to automatically collect data on face-to-face communication between collocated members of the teams. A sociometric badge is a device containing a microphone, an infrared sensor, accelerometer, and radio transceiver.

When studying communication, the badges were used in two ways. First, each team members wear a badge around his or her neck. We collected data on both physical proximity between team members, as well as their communication behavior; who talks with whom, how and when. Secondly, badges are attached to physical locations, such as task boards, workstations, and other interesting places, such as cafeterias and meeting rooms. This helps us understand where the communication takes place.

Relating face-to-face communication with computer mediated communication allowed us to further measure the total amount of communication in ASD projects, to understand whether mediated communication could act as a substitute for face-to-face communication, and how face-to-face communication patterns and behavior differed from the electronic media use.

3) Interviews: By interviewing project team members — including project manager, scrum masters and developers — we gained insights on communication in the project studied: what were the benefits and challenges in using email and IM in the project, and what are the perceived experiences and attitudes towards face-to-face, email and IM use in the project. Additionally, interviews allowed us to understand the project goals, software development process and practices used, the product being developed, and about the general organizational context of the project.

4) Observations: We have also used non-participative observation in selected situations, such as meetings and planning sessions as well as normal daily work, to gain qualitative information and understanding about the communicational and group dynamic structures of the team, and about the agile process used in the project. Observations are also used to discover possible contradictions between the collected data and actual team behavior.

5) Feedback sessions: We arranged a feedback session with the team. In the feedback session, we presented our results to the team members, and discussed about the results and our interpretations in order to validate our understanding and findings.
C. Data analysis

The communication data collected from the case projects was analyzed quantitatively. We extracted communication frequency — how often people send email or instant messages to each other, or talk face-to-face — and the amount of communication — number of messages via email and IM and number of minutes of face-to-face discussion. The analysis of communication frequency and amount was analyzed on daily level, by aggregating all communication by the whole project team or site-specific subteams during the day together.

Face-to-face communication was analyzed based on data collected with sociometric badges. A sociometric badge recorded the amplitude of speech every minute, and we used a specific threshold to detect speech based on amplitude. Samples above a given threshold were considered as speech time, and samples below the threshold were considered non-speech time. Same threshold value was used for all team members. By aggregating the number of samples above the threshold, we came up with the total number of minutes per day the team was involved in face-to-face communication.

In electronic communication — email and IM — we used TO and CC fields in individual messages to determine the participants of communication. For face-to-face communication patterns, we used infrared sensor on sociometric badge to detect the face-to-face links between team members. We used social network analysis to extract both network metrics (such as network density, centrality and hierarchy measures [5]) and communication patterns from communication networks collected from both face-to-face interaction and electronic communication. With social network analysis, we were able to visualize the communication patterns via different communication media, by considering both links between individual team members and the frequency of communication between team members.

IV. Results

In the following chapters, we present our results on communication amount and patterns in the project studied.

A. Verbal communication

The face-to-face social network structure is visualized in Figure 1. During the period of study, the team members located at different sites did not meet each other physically, i.e. team members did not visit other sites. Therefore, the team members in different sites are not connected in the face-to-face communication network, as they were not communicating face-to-face with each other.

Within sites, we can see fairly well connected face-to-face communication networks. For Czech site, all team members are connected, representing the fact they were working in a same shared office room, and met each other constantly. From the social network, we can further see, that some team members in Norway were working remotely, and thus were not communicating face-to-face, even while they were formally working at the same site. Beyond the people working remotely, the face-to-face communication was tight and equal, as can be predicted from the Scrum practices used by the team; the team had daily Scrum meetings, where all team members were supposed to be present, either physically or remotely via teleconferencing.

Figure 2 presents the time the team collectively used in verbal communication per day. Verbal communication amounts here represents only talking time — time a person was talking, so e.g. in meetings, we count only one minute for the whole team, if only one person is talking at the time.

The amount of verbal communication per day had a large variance between different days during the study. This variance suggests that there may be fine-grained structure of work in agile software development projects beyond the sprint level: different tasks require different amount of communication, and these tasks are not evenly distributed within one sprint. Especially we noticed higher peaks in communication at the end of the sprint, which can be accounted for sprint demonstration and planning activities.

Also at the end of the data collection period, we see rapid decrease of the amount of verbal communication, which is due to sliding ramp-down of data collection.

B. Email communication

Email was a communication tool used actively by all team members. The following email communication analysis is based on contents of six team members, three from Czech site and three from Norwegian site.

The communication network based on email (see Figure 3) visualizes the usage pattern of email in this project. From the network graph we can see the connections between individual team members from different sites. A connection between individual nodes (circles) show that messages were sent from one person to another, and the length of the link represents the amount of messages sent between those individuals: shorter distance between nodes means more messages between those individuals.

Almost all Norwegian and Czech team members are connected to each other, thus the density of the email commu-
The communication network is high. There are no site-specific cliques — set of individuals who are tightly interconnected but loosely connected to others — in the email network, which means that email communication patterns do not significantly differ depending on the site the person is located in.

The total amount of messages sent and received by sites in the project studied are shown in Table I. The first part of the table shows the message counts sent to primary recipient (TO-field in email) and second part shows the secondary recipients (CC-field in email) for messages sent from respective sites. In this table, the same message is counted multiple times, if it had multiple recipients.

The reported main usage for email — communicating team-wide, non-urgent issues — is reflected also in the average number of recipients per message. For primary recipient (TO-field), the average number of recipients was 6.2 per message, and the average of secondary (carbon copy) recipients was 1.01 per message.

There was also some use for hidden recipient field (BCC, blind carbon copy), but as its perceived usage was relatively rare as well as accurate measurements are challenging due to information hiding, we have excluded blind carbon copies from this analysis.

Figure 4 present the number of email messages sent in total for each day. From the graph we see that the volume of email messages per day was on average around 30−40 messages per day. In comparison to face-to-face communication amount on daily basis, we see very little variation regarding different sprint phases: the number of email messages per day is fairly constant during the whole period of study.

**C. Instant messaging**

The social network graph for instant messaging is presented in Figure 5. As with email network graph, a link between nodes indicates that these two individuals have participated the same instant messaging session, and the distance between nodes represent the amount of messages between these two individuals.
In comparison to email network, IM network graph shows more distinction based on the location of team members. We can clearly see a clique of Czech team members, who are tightly connected to each other, which indicates that they actively use IM to communicate with each other. Two Czech team members never communicated via IM outside their site, while the other two Czech team members were also communicating via IM with remote team members in Norway and in Finland.

When analyzing the communication patterns between Czech and Norwegian sites, we see a strong evidence of such "bridgehead" role. There are two team members at the Czech site who are communicating closely with one Norwegian team member over instant messaging. The volume of this communication seems quite high, as Czech team members in general communicate a little bit more with Norwegian team members than with the team members on their own site (see Table II).

<table>
<thead>
<tr>
<th></th>
<th>Czech</th>
<th>Norway</th>
<th>Finland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech</td>
<td>820</td>
<td>846</td>
<td>262</td>
</tr>
<tr>
<td>Norway</td>
<td>846</td>
<td>2528</td>
<td>112</td>
</tr>
<tr>
<td>Finland</td>
<td>381</td>
<td>172</td>
<td></td>
</tr>
</tbody>
</table>

Interestingly, we see tighter connections between Finnish and Czech team members in instant messaging, than between Finnish and Norwegian team members. The number of messages sent from Finnish team members to Czech team members is indeed higher than the number of messages sent between Finnish and Norwegian teams.

Figure 6 shows the number of IM messages exchanged between team members per each day. The variation of messages from one day to another is higher than daily variation in email communication volumes. We also found out that most IM discussions involved more than 2 participants: group chat functionality was often used to quickly resolve issues between site subteams, and also to provide awareness for other team members about the ongoing, non-verbal discussion.

V. DISCUSSION

The average time used for communication per team member per day was 45 minutes for the studied project. This number is significantly smaller than the amount of verbal communication what earlier research [20] has reported for software development teams — 75 minutes — and the measured value includes not only unplanned, informal communication, but also formal meetings, which include at least 15 minute daily Scrum meeting every day.

We also found out that the amount of verbal communication within the team varies heavily from one day to another, indicating either differences in communication needs, or differences in opportunities for verbal communication between different days. At this stage of our analysis, we are not able to conclude whether the cause of this variation comes from the communication need, team members’ availability for verbal communication, or both. However, we do see similar variation in instant messaging communication, and no variation in email communication, which suggests that at least communication need could be major contributor to this variation.

On average, we observed about 80–100 IM messages sent daily between team members. At this stage of analysis, we are not able to estimate the time team members were actively using IM for communication. However, based on previous work and our observations, team members were often multitasking while active in IM: they used IM to ask clarifications and simple questions from other team members, and while waiting for reply, they kept working on other tasks. This makes it fairly difficult to estimate the actual time spent in IM communication.

As IM messages are typically composed of single sentence (e.g. a question or short answer to that), the amount of information conveyed within one message is considerably smaller than that of email messages. Also email messages are intended to be more stand-alone and less sensitive of the context of communication, and thus the preparation and producing email messages will take more effort than single IM message. As we were not able to record the actual use of email and IM clients in the team members’ workstations, we do not have exact values for the time spent exclusively on IM or email communication.

The average amount of email messages per day was 40–50.
In the project studied, the main use for email was to communicate team-wide issues. This use of email was reflected both in the social network graph based on email messaging — shows all team members from all sites equally immersed in the graph — and in the average number of recipients in each message — indicating that many email messages were sent to all project team members.

Furthermore, the site-spanning use of email can be seen when counting the messages sent and received between the sites. From the message counts, we can see that for Finnish and Czech sites, the majority of messages were sent across sites: email was used especially for communication between the Norwegian and Czech sites. However, Norwegian site used email even more in communicating internally than to communicate with other sites. This may be affected by both the specific roles assigned to Norwegian team members (project manager, scrum master) as well as the frequency of remote working at that site.

The use of instant messaging was more targeted towards communicating with team members at the same site, while site-spanning communication was still somewhat used. It seems that the communication across sites was more focused on few team members (“bridgeheads”), who then supposedly relayed information to other team members located at the same site they were located in.

A. Limitations

Our study reported in this paper was limited to a single case project only. This severely limits the potential for generalizing the results beyond the context of the project studied. However, we used several methods of inquiry, and our findings were in line with the interviews conducted with the team members, as well as the feedback session, where we presented our findings to the team members and discussed about them.

One major limitation in our study on verbal communication is related to the method of data collection. We used novel method — sociometric badges — for capturing the face-to-face interactions between team members, and while this data collection method has been used before (see e.g. [6], [23]), it still must be considered as an experimental way of collecting such data. However, based on our observations of daily work within the teams, we believe our measurements of the daily face-to-face communications seem not to be significantly wrong. On the contrary, we believe that the amount of verbal communication can actually be smaller in this case than in cases studied in earlier research, as some of the communication was conducted via mediated communication media, such as email and instant messaging, potentially reducing both the need and possibility for face-to-face discussions.

For the electronic communication, we were able to collect only email and IM logs from Norwegian and Czech team members: we have no data from Finnish team members, but we are able to infer their communication with Norwegian and Czech team members by looking at the messages received by these sites. However, we are not able to see how much communication did take place within Finnish site. This limitation is however minor, as the roles in Finnish site were mostly managerial, and we assume most of the project related communication from Finnish team members did take place across sites.

VI. Conclusions

We found differences in the daily variation of communication amounts across different media as well as different network structure of communication with different media.

As an answer to research question 1 (How much communication takes place face-to-face, via email and via IM in distributed ASD projects?), we conclude that the amount of face-to-face communication in the project studied was on average 45 minutes per day, an amount that is lower than earlier studies [20] have reported. While this study has limitations related to assessing the amount of verbal communication per day, we believe that in distributed software projects, the amount of verbal communication is lower than in collocated projects, as mediated communication, especially IM, must be used to communicate across sites. We also found out that both email and IM was used actively on daily basis, with average of about 40 email messages and about 80–100 IM messages per day.

For research question 2 (...), we found out that the communication networks differ radically between face-to-face, email and IM communication. In distributed projects, face-to-face communication is obviously limited for intra-site interaction only, but surprisingly also IM network shows tendency to communicate only with the people on the same site, having specific “bridgehead” roles for communicating across sites. On the other hand, in email communication we did not see such
distinction between sites. This supports our earlier findings that email is often used for inter-site team-wide communication, while IM was used more for quick questions and answers between people who are working on the same tasks [15]. Additionally, by looking the patterns of communication amount variation between different days, we found some evidence that IM is used as a replacement for face-to-face discussion.

A. Future work

This paper presents a preliminary study on communication amount and patterns using face-to-face, email and instant messaging in a single distributed agile software development project. The analysis of the data as well as the development of analysis methods is still work-in-progress, and for this paper we were able only to disseminate our first impressions of the potential findings in the data. Especially we have now aggregated all data on team level, and while agile teams should be cross-functional and composed of generalists, we believe that more detailed analysis of communication patterns and tool use differences between individuals and different roles, such as product owner, scrum master and developer, would be interesting and beneficial for communication improvement in distributed agile teams.

In this paper, we have also presented our findings based on communication data only. Based on this data, we are not able to conclude whether the observed communicational behavior is beneficial for team members or the project goals, and how the communication should be organized for better project performance and job satisfaction. Future studies should address the effect of media choice and project performance, by assessing the relationship of given communication patterns and project outcomes.

REFERENCES


