# The Systems of Instant Messaging - Draft - TO BE REPLACED BY FINAL ON 12/6

Devashish Purandare
*University of California, Santa Cruz*

Submission Type: Survey

## Abstract

come up wid a gud un

## 1 Introduction

Instant messaging (IM) is not a new technology, it started in 1792 with the invention of the telegraph. And while the core principles remain the same, each year, more and more things are made possible with IM, the complexity grows and the scale too. There are 4.7 billion cell phone users worldwide[1]. This explosion in cell phone usage, coupled with improvement in network coverage with high speed data access worldwide has caused instant messaging to grow on a massive scale. Several popular instant messaging apps boast of more than a billion unique users.

This scale presents several challenges which are unlike what are faced by most distributed systems. Each client is unique and the connectivity to each client is unreliable. There could be multiple instances of the same client on different architecture. While consistent ordering is not a compulsion, there's a need to maintain at least causal consistency. With clients located at geographically different locations, in different timezones, the need for security and privacy, the challenge presented by modern instant messaging systems becomes more complex.

## 2 Motivation

While a lot of work is work has been done on developing instant messaging services, little has been done on aggregation of all the techniques into a single cohesive compilation. This paper aims to trace the history of instant messaging, the design decisions that went into the instant messaging technique of each era, the sudden growth of IM, and the future. The paper will focus on 5 main areas, Chronological time-line of IM, the challenges each protocol faced and the way those were addressed, what is the state of the art and new technologies in instant messaging, the modern extensions to IM, Privacy and Security in IM. The aim is to create a definitive document tracing the history of IM and the modern techniques.

## 3 Organization of the paper :

The paper is organized chiefly as follows, section 4 talks about what is an instant message and the challenges faced, section 5 will focus on historical systems, section 6 will address the arrival of internet and subsequent systems and section 7 will be about current systems. The paper will be characterized by two main events which gave a boost to messaging, namely the arrival of Internet and the arrival of Smartphones. While internet created a world-wide network of connectivity to allow for instant communication, smartphones helped spread internet, with internet connectivity reaching every person as well as lowering the technological barrier for using IM.

## 4 Instant Messaging

To talk about IM, it is important to first define what IM is. It is tempting to define IM as instant transmission of short messages, but there need not be any limit to message length. If we limit IM definition to P2P (point to point) communication, group messaging and chat rooms will be left out of it. And if we limit the definition just to text messages, modern systems will not be taken into account. So to make our definition as broad as possible, we define IM as

*"Communication over a bounded time"*

# 5 Before the World Wide Web

The need for instant communication has been around ever since the dawn of civilization. The earliest forms of communication over a distance known to us is smoke signals. Smoke signals were used by the Greeks to communicate in wars, and as we will see, the need to communicate in wars has given birth to a lot of systems of communications as well as some of the most fundamental problems faced by distributed systems.[2] The first succesful semaphore telegraph network was operated in France as early as 1793. While the early Telegraph systems used optical communications, with the arrival of electricity, electric telegraphs, using the Morse code became the first major worldwide communication system. The research on Hertz-waves gave rise to radio communication, allowing wireless transport of communication. Telegraph lines were laid under the sea in the first ever world-wide network for instant messaging. [3]

## 5.1 The Electric Telegraph

The way electric telegraphs worked was, much like telephone line, they requested a connection to the center, a switching center would set up a session and a line between them, and data could be transmitted over it. The switching did not allow for communications to a node when it was a part of another session. Because of the centralized nature of Telegraph system and the low number of messages, this was a feasible and practical solution for Telegraph. Early communications over phone line, such as dial up internet and fax used the same technique.

It might seem tangential to include telegraph in this survey, but Telegraph gives us surprising similarities and for much of its lifetime, faced the same challenges that messaging today faces and gave rise to many of the methods that are used widely today.

As Telegrapgh gained popularity, the centralized switching network soon began to create hotspots at hubs like London which led to serious slowdown due to congestion. The way Telegraph companies got around this was through an ingenious technique – setting up steam powered pneumatic lines, which forwarded messages to nearby centres, who then processed and sent the message over the telegraph line. This works much like the modern load balancing techniques used in distributed systems.

Each message had a header space, where it was "Timestamped" by the receiver with the source and destination address, and by the time it reached destination, it had the complete history of the path it had been routed through. A century later, email would use the exact same technique. The Telegram centers which spoke frequently, set up a Nickname ' system to reduce the size of this header.

The transmission of messages, which involved human operators reading and retransmitting the message, gave rise to codebooks, ciphers and encryption, not too dissimilar from the techniques that are used today. Paul Reuters (founder of the Reuters agency) famously used 3 homing pigeons with the same message to guarantee reliable delivery. The cost being dependent on the length of message gave rise to techniques for data compression.

There are many parallels between the uses of Telegraph and the current messaging systems. Telegraphs were famously used for playing games, transferring money, exchange of information and even for 'online' marriage services. The operators at telegraph decoding stations would use the service for informal meetups and chats after working hours were completed. [4]

As we move along the time, we will see, much of the progress in messaging, is just reinventions to adapt the same technology to a newer medium of information flow. With telegraph and telephone, it was electricity, with email and IRC it was Internet, and with modern messaging systems, it is smartphones.

### 5.1.1 The Radio Facsimile (FAX)

## 5.2 The Telephone

The discovery of Telephone was no accident. Alexander Graham Bell, working on a harmonic telegraph, a device which used frequency of sound to transmit messages, it would be possible to transmit sound over electrical lines directly, reducing the need for the message to be decoded at the other end. The first working Telephone in fact was marketed as a speaking Telegraph machine which required to manual interpretation. Telephone became an instant, building on available telegraph infrastructure, it largely led to the decline of Telegraph.

The telephone followed a very similar path to the Telegraph, and its development was led by many of the same people who worked on Telegraph systems - Thomas Edison. From transcontinental telephone lines being set up to a trans-Atlantic line, it

followed the same pattern as Telegraph, and the internet would again follow. The biggest achievement of the Telephone was that it managed to free communication carried by and developed by mostly postal and railway services to something which private citizens would own.

Much of the research on reducing congestion in Telegraph was carried over to telephones, with duplexing and quadraplexing to boost bandwidth and modulation to reduce the amount of data to be transmitted.

The invention of Hertzian Radio waves, which could be used to transmit sound over short distances, soon gave way for wireless radio communication, typically over short distances. This led to the invention of one of the biggest catalysts for instant messaging development, the mobile phone.

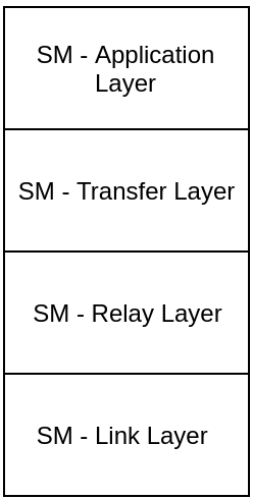


Figure 1: The architecture of SMS

## 5.3 The Mobile Telephone and Wireless Communication

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ac quam quam. Sed convallis mollis magna ut molestie. Quisque convallis, justo at blandit dapibus, est erat hendrerit est, sollicitudin sodales felis odio cursus tellus. Aliquam ultricies non tellus in accumsan. Fusce lobortis tortor et enim posuere, ut tincidunt ipsum interdum. Nunc a cursus metus, ut porta arcu. Aenean non egestas arcu, in fermentum enim. Curabitur nisl nunc, rutrum ut enim non, lacinia placerat arcu.

Sed nec odio quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Duis laoreet non urna nec pulvinar. Mauris sit amet semper massa. Donec blandit massa nec nunc malesuada imperdiet. Praesent ac euismod sapien. Fusce mattis felis est, ac rutrum ligula semper id. Suspendisse rhoncus ex in mauris viverra ultrices. Donec quis enim scelerisque, interdum dolor sed, viverra felis. Duis ut magna id diam ornare posuere.

## 5.4 The Short Message Service

### 5.4.1 The Technology

In the year 1984, while working with GSM networks, Friedhelm Hillebrand and Bernard Ghillebaert realized that, the control signal mechanism to diagnose and configure GSM could be used to relay short messages over mobile networks. The medium, optimized for telephone put restrictions on the payload such that only 160 characters of 7 bits each could be transmitted. SMS was one of the first open platforms for application and vendor independent messaging. While initially introduced for GSM networks, it was soon ported to work with GPRS and CDMA networks. The big appeal of SMS was its global compatibility, SMS could be sent over virtually every network worldwide with just a software update to the existing infrastructure.[citation needed]

SMS allowed peer to peer communication, passed through a wireless carrier, with each message limited to 160 characters.
Terminology :

- Short Message Entity - SME - application that sends and recieves SMS
- Service Centre - SC/SMSC - route, store and forward messages.

The SMS protocol has a set of features : SMS can get and generate delivery reports, add a reply path, execute an application and act as a command. Due of the nature of mobile networks, particularly 2G networks, the SMS protocol allows store and forward messaging. When a SME sends a message, the SMSC keeps a copy of the message and forwards it to the next base station. This goes on until the message reaches the station that has access to the receiving SME. If the device is not connected to the network, the SMSC holds the message for a specific amount of time waiting for it to reconnect before discarding the message. This specific amount of time can either

be specified by the message originator, but to simplify communication, most companies use a validity period of 48 hours (2 days).[5]

SMS has a big limitation that it is restricted to a size of 160 characters * 7 bits. SMS uses UCS2[6] unicode's precursor to UTF-16, a character set which allowed complex characters such as Chinese and Arabic Script encoded in a small size. However this makes it particularly limited when sending messages in complex scripts which requre more space - such as indic scripts or emoji, limiting the characters to 72 or lower. While there was disagreement on the length of SMS, concatenating multiple SMSs allowed operators to offer greater character limits. Concatenating was introduced later[7], where carriers used multiple messages with sequence numbers and metadata to send a bigger SMS.

#### Application Layer

This includes all the applications that make use of SMS inlcuding games, banking, rental, value added services.

#### Transport Layer

Message is encoded in a sequence of octets and made ready for transmission with details such as return address.

#### Relay Layer

The relay layer is responsible for the 'store and forward' routing of messages.

#### Link Layer

Link layer is the actual physical medium over which messages are transported.

### 5.4.2 Impact of SMS

SMS gained immense popularity because of its global compatibility and interoperability. The low cost of sending a message, the ability to receive a message even in adverse network conditions, the simplicity of sending a text and the low barrier of entry made SMS the most successful forms of instant communications in the world. It is estimated that 6.1 trillion messages were sent in the year 2010, that is about 200,000 messages every second 5.3 billion subscribers of cellular networks existed and wireless networks have reached where wired networks could not [8]. This connected billions of people worldwide and allowed for a cheap and easy way of communication.

SMS spawned several products, directly or indirectly which had great impact on the world. Due to the size limitation on early cell phones, most phones had a 12 key keypad which made it very hard to type. Several algorithms were invented, such as T9, to allow predictive typing which would make it much easier to type of cell phones. Much like the telegraph, SMS length limitations soon spawned a language of there own, simplifying and shortening many words to make typing them faster and sending them easier. The human emotion is not limited by words, soon using expressions like `:)` to denote a smile or `<3` to denote a heart became popular. As the popularity of ASCII emotion icons (emoticons) grew, companies began to parse and integrate animated, colored or custom versions of emoticons in their applications. To distinguish their service, Japanese network carriers started offering custom emoticons (Emojis) to users. Noticing the popularity of emoticons, unicode started adding emojis in an effort to standardize offerings. Unicode now offers over 2000 emojis, with options for skin tone, gender and professions [9].

In 2015 "😂" was declared as the word of the year by Oxford dictionary, noting it was the most searched word that year. [10]. As with telegraph, an industry soon appeared around SMS, SMS stock updates, sports updates, offers, SMS payments, SMS verification. SMS allowed two-factor authentication, a security measure that would make evrything much more secure and tedious. The 160 character limit of SMS was adopted by Twitter, limiting tweets to 140 characters. The idea of limited text instead of being restrictive spawned a new generation of social interaction, where volume and formality don't matter. Many companies modified their services so that they could be completely accessed using only SMS. Today, every phone plan in the US comes with unlimited text messages. GSMA has decided to introduce RCS, a protocol discussed in detailed later in the work to replace SMS.

## 5.5 Enhanced Messaging Service and Multimedia Messaging Service

### 5.5.1 Enhanced Messaging Service

Enhanced Messaging Service (EMS) was introduced as an improvement to SMS. It was backwards compatible, so it could fall back to SMS if the device

did not have EMS capabilities. EMS was a joint effort by several mobile phone manufacturers and it was completely implemented in the application layer of SMS so that the data was sent over SMS. EMS added several capabilities to SMS, adding the ability to send and receive bitmaps. EMS also added rich text formatting capabilities such as making text **bold** or *Italic*, changing the font, changing the font color and size. EMS also allowed the capability to send sounds, either predefined (where the actual tone was not transmitted) or custom. Custom sounds used the iMelody format to specify notes, composer and other metadata, allowing additional operations such as controlling lights and vibrations [11]. EMS also allowed specifying 4 8x8 or 2 16x16 images to serve as an animation.

EMS was further improved upon to Extended-EMS. Unlike EMS, it was not backwards compatible, but it extended several capabilities of EMS by adding options to send bigger images, 64 bit color images, polyphonic ringtones and vector images. While Extended EMS offered a lot of versatility, the underlying medium was still SMS.

#### 5.5.2 Multimedia Messaging Service

With improvement of networks to 2G and 3G, transmitting data became easier, faster and more reliable. There was a need to ensure interoperability between the internet and phones. SMS and EMS were defined completely in the bounds of the network architecture and development of MMS required a lot of standardization, mainly with Wireless Application Protocol [12] and internet protocols. MMS added interoperability between SMS and email. Development of MMS required extensive collaboration between 3GPP which defined the content structure and format of MMS and WAP forum which adapted it for the internet. MMS added several important capabilities to instant messaging : the ability to send pictures and videos, the ability to send documents and email and the capability of voicemail.

In MMSE (MMS Environment) the MMS server stores all the messages which are sent over MMS and the subscriber gets a notification when a message is sent to him. The subscriber can then retrieve the message from what is stored on the MMS server message store. The MMSE includes several components and interfaces :

- MM1 - Interface between the user and MMSC (Multimedia Message Service Centre).
- MM2 - Interface between MMS Relay and MMS Server.
- MM3 - Interface between MMSC and external servers - email/SMS
- MM4 - Interface between two MMSCs for forwarding messages.
- MM5 - Interface between MMSC and routing information.
- MM6 - Interface between MMSC and User databases.
- MM7 - Interface between MMSC and Value Added Services - VAS.
- MM8 - Interface between MMSC and billing services.

Out of these MM2, MM6 and MM8 were never implemented or standardized. MMS was transmitted over TCP with a WAP proxy on top allowing interoperability. Hence MMS routing followed the handshake, data security and reliability guarantees offered by TCP.

MMS followed the specifications of [RFC822] [13] and [RFC 2822] [14] which were defined for email and later adapted to MMS. MMS followed and gave rise to the popularity of several W3C standards such as XML, SGML, HTML, XHTML, WML. MMS are routed through two different methods either by email eg. `user@mms.serviceprovider.com` or via Multimedia SISDN with unique identifying numbers.

MMS works in a similar way to SMS however it makes certain modifications for delivery. Because network connectivity is not guaranteed to be fast, the user just receives a notification for a MMS and can defer the retrieval of the actual message later. Deferring retrieval makes the MMS server store the messages for a longer amount of time. For the files which are too large, The Real Time Transport Protocol (RTP) and the Real Time Streaming Protocols (RTSP) were used to stream data bit by bit to the device [15, 16]. RTP and RTSP use time stamps and sequence numbers to order data and provide streaming.

While it did not gain the popularity that SMS did, MMS gave rise to many important standards and ways of communication of data. MMS gave rise to concerns of sharing copyrighted materials and gave rise to standards of Digital Rights Management

(DRM). MMS also gave a boost to various established standards of data exchanges such as SOAP and XML. Most importantly, MMS bridged the gap between the Internet standards and Cell phone standards allowing for easy exchange of data.

## 5.6 ARPANET and the dawn of the Internet

Advanced Research Projects Agency Network (ARPANET) began as a research project in a joint initiative by the US department of defense and universities across the US. It was the first packet switched network implementing the TCP/IP protocol on such a large scale. Soon the ability to send message was added to ARPANET via various RFCs [17, 18]. Ray Tomlinson, regarded as the inventor of email used the "@" operator to differentiate between user and host name in the address. This was further extended to become `user@host.domain`, the standard email address formats that are widely used these days.

## 5.7 Other technologies :

Various other technologies existed for instant transmission of messages, such as messaging over Bluetooth in a peer-to-peer local network. Various technologies also allow sending messages over infrared signals, or through a local network such as LAN or WiFi.

# 6 The World Wide Web

The Internet, due to its scale, nature and openness soon became the ultimate medium for transport of messages. Much like telegraph, it became a way to transport data across the world at an unprecedented speed and a very low cost. This gave rise to the acceptance of instant messaging as form of formal communication and it became the de-facto medium to do so.

## 6.1 Email

## 6.2 Chatrooms

The rise of popularity of Internet and its widespread use gave rise to message boards for people interested in a certain topic. Message boards would be websites, discussion forums, news and marketplaces. Soon websites started allowing users to participate in real-time conversations giving rise to "chat rooms."

### 6.2.1 Internet Relay Chat (IRC)

IRC was created as an open text-only messaging protocol which would use TCP to connect to IRC servers creating an IRC network. IRC does not specify the text encoding and the RFCs [19, 20, 21, 22, 23] used for reference are rarely used which gave rise to many issues initially.

IRC Server protocol [22] pings all its connections periodically polling them, if a connection is unresponsive, a termination procedure is followed. Connection of two servers is a critical and error prone area in IRC. IRC was designed for slow networks, often dial up networks over telephone lines. To speed up data transmission, IRC servers support compression of data streams being sent. However this causes further problems in connection and forwarding of messages since all servers in a network, particularly external servers may not support compression.

To ensure consistency, IRC uses state information, mainly the information about the states of clients, servers and channels. This state machine information can then be used to determine where the connections should be cut off as IRC can only work in acyclic graph formation and formation of cycles is considered a collision. Collision can be because of two servers connecting to each other despite there being a path between them or because server nicknames turn out to be the same in a large enough network. When a server is terminated, the server sends a SQUIT message to all other servers to ensure that the network remains consistent.

Another measure that the IRC protocol takes is that each server maintains a record of all the recent nicknames so that in case of a server split or nickname change race condition, the network would continue to function and have information to resolve the conflict.

IRC also offers a way to deal with rogue or malicious client flooding the servers with messages by keeping a track of timestamps of the messages. After a certain threshold, each message from the client is penalized with a time delay (initially 2 seconds) for each messages, the delay can be incremented with time.

IRC had many problems from start such different standards of implementation causing interoperability issues. IRC is not very scalable fundamentally due to the requirement that each server should know

about every other server and network can only be an acyclic graph. This also presents several privacy issues. IRC allows the users to pick labels for the nickname, channel name and service name. With no duplicates allowed, this often results in collisions. The channel and server lookup causes great problems with scaling with implementation algorithm having complexity $O(n^2)$ .

IRC became very popular with over 1 million unique users every day, but its popularity was restricted to certain demographics such as developer communications, inter company collaboration or discussions about a topic. While it never gained the popularity enjoyed by the likes of SMS, it gave rise to collaborative messaging and data interchange apps like Slack and Yammer which took the core concept and implemented it with modern technologies. As of 2016, IRCv3 is in works, adding protocols for user presence, file, image, audio and video exchange and adapting IRC for modern networks.

## 6.3 eXtensible Messaging and Presence Protocol (XMPP)

XMPP was originally open sourced as Jabber protocol and is one of the most widely used protocols for instant messageing. As its name suggest XMPP is based on XML format. XMPP is a part of open standards and is maintained by the XMPP work group a part of IETF.

XMPP was defined as an open internet standard with RFC6120 [24] in 2004 and was developed upon to create an open standard for messaging and presence.

### 6.3.1 Presence

Presence allows clients to subscribe to and enquire about the status of a particular client. This allows messaging applications to provide richer experience to users, showing the other user as "online", "busy", "away." This can be a very useful feature for users, and helps improve communication by providing additional information.

### 6.3.2 IQ

IQ in XMPP stands for Info/Query which is useful for setting up connections, getting the status of the network/service, and troubleshooting.

```
<stream>
<presence>
    <value/>
</presence>
<message to='recepient'>
    <body/>
</message>
<iq to='server'>
    <query/>
</iq>
...
</stream>
```

Figure 2: XMPP stream

## 6.4 Messaging with XMPP

All the exchange of information in XMPP happens in XML streams. XMPP defines two types of XML communications : XML streams and XML stanzas. XMPP also allows definition of a gateway to other protocols which would perform the translation. This has allowed many messaging applications to offer support or partial support for XML. Clients connect to XMPP servers over TCP, here multiple resources (such as multiple devices) can connect with the same client id to the XMPP servers on behalf of the same client. This is a very important feature, considering the fact that it allows users to use messaging over different devices like laptops and phones at the same time.

Messaging involves asynchronus exchange of data with relatively small payloads. To work with this XMPP defines XML Streams and XML stanzas :

### 6.4.1 XML Streams

The XML stream starts with a `<stream>` tag and can send an unbounded number of XML elements between this tag and the `</stream>` tag, which denotes the end of the stream. A response stream can be set up immediately to send responses.

### 6.4.2 XML Stanzas

An XML stanza is the first child of the root `<stream>` and can contain several sub tags and attributes. It is used for the presence, iq, and other features offered by XMPP.

The structure of XMPP data exchange is usually something like Figure 2 [24]

XMPP allows encryption of streams using TLS or SASL and requires response stream to have a different key than the stream it responds to. However the server should ensure that the same scheme is supported by the receiving server.

If the receiver detects an error it may send with an `<error/>` response specifying the error, errors in XMPP are considered unrecoverable and the data needs to be retransmitted. XMPP by default uses UTF-8 making it easy to have a set encoding style which is useful globally. XMPP manages a roster of client's contacts and allows only a user-approved subset of the contacts to subscribe to presence data [25]. For security and privacy, XMPP also maintains privacy lists, allowing users to block other users or communications.

XMPP battles with forging addresses in its Server-to-Server authentication protocol which stamps 'from' and 'to' at every step of communication. However it is still possible to forge address in this system by using a malicious server or by attacking the DNS. XMPP also suffers from the address mimicking flaw that email suffers from, partially due to the use of UTF-8 [26].

## 6.5 Messengers

The 1990s and early 2000s soon became the golden age of desktop based instant messaging with dozens of competing messengers fighting for a share of the market.

### 6.5.1 AOL Instant Messenger

AIM was at one time the most widely used messaging app. AIM used the OSCAR (Open System for CommunicAtion in Realtime) which AOL developed. Despite its name, OSCAR was a proprietary protocol, and AOL went great lengths to keep the protocol from being used by competitors [27].

AIM 'running man' icon became one of the most recognized the symbol of early 2000s, however its popularity fell sharply with the rise of social networking sites and Google-talk. AOL tried porting AIM to XMPP, but the effort was abandoned later. AIM also supported sending small files, audio-videos and real time games.

### 6.5.2 ICQ

ICQ (read I seek you) was released as an open protocol for messaging in 1996. ICQ quickly gained popularity with their method of assigning Unique Identification Numbers (UINs) to users allowing them to share them for connections. Blackberry messenger and Snapchat would later follow this model.

### 6.5.3 Yahoo Messenger

Yahoo introduced its own messenger and accompanying protocol for instant messaging. The protocol would include presence, file sharing, gaming and ability to send stickers and have custom avatars.

### 6.5.4 Google Talk

Google introduced an implementation of XMPP known as Google-talk alongside their email. Google Talk had all the features included in XMPP such as presence and multi-device support and some extensions such as the ability to send images, and files. With the launch of Google+, Google later replaced this protocol with its own proprietary Hangouts protocol.

### 6.5.5 Skype

Skype introduced one of the first server-less Peer-to-Peer messaging and video-audio calling protocol. This protocol was very successful initially, as it ensured smooth communication even with poor network connectivity because there was no server connection involved, allowing optimum utilization of the bandwidth. However, after acquisition by Microsoft, the protocol was converted into a server based protocol because it would not work well on smartphones, where battery life and processing power is constrained [28].

### 6.5.6 MSN Messenger

MSN Messenger added several features to messaging such as social network integration, ability to send and view albums, offline messaging. It was one of the first messengers to be available for the mobile platforms. It was later discontinued when Microsoft acquired Skype.

# 7 The Smartphone Era

In the last few years, the processing power and connectivity of phones has improved exponentially. Better hardware and more capable software to accompany turned the cell phone into a smart phone - a

powerful device with abilities comparable to a computer. Improvements such as 3G and 4G as well as fast WiFi networks have improved the connectivity to phones significantly. With more and more people using smart phones, instant messaging got a huge push with millions of people experiencing the Internet for the first time. This growth was particularly high in countries such as India and China, which had poor network connectivity and awareness before low cost smartphones flooded the markets.

It is no surprise therefore that Instant Messaging is the biggest it ever was with trillions of messages sent every year, hundreds of thousands of them every second.

## 7.1 Social Networking

Smartphones gave a helping hand to the already growing social networks and the most popular social network - Facebook has more than a billion unique users. Twitter has millions of users who use it everyday. These networks added instant messaging abilities from day 1, including ability to send one-to-one personal messages.

Facebook introduced the ability to chat from the first version of the website. This ability was soon bundled into a standalone product - Messenger for smartphones. Facebook Messenger uses the MQTT protocol : see 7.2. To improve user experience Facebook made several modifications to the messenger, including ability to send SMS and continue the conversation if user is not connected to the internet, and releasing an optimized, light version of the app for phones with low memory, processing power and for areas with poor connectivity [29].

Facebook then acquired Whatsapp see 7.4 to bolster its smartphone offerings and has built a platform around the messenger for people to develop on allowing chat-bots, games, transactions over IM [30].

## 7.2 Message Queue Telemetry Transport (MQTT)

MQTT is a lightweight protocol, specially designed for constrained environment which provides lossless bidirectional ordered communication of messages. MQTT is an open standard, defined as ISO/IEC 20922:2016 [31]. MQTT uses websockets and other web based protocols such as WebRTC for communication. It is optimized for small payloads and delivery constraints [32]. MQTT gives the application control of message delivery allowing 3 different mechanisms using different methods :

1. At most once - Messages are sent only once, and message loss may occur.

2. At least once - Messages can be resent and received multiple times, but all messages are delivered.

3. Exactly once - used for transactions.

## 7.3 Smartphone only Messengers

With the growth of Smartphones, messengers designed for smartphones first arrived with special features designed and optimized for the use with smartphones.

## 7.4 WhatsApp

The most successful of these has been WhatsApp, with over 1 billion users as of 2016 and 42 billion messages sent every day [33]. WhatsApp uses FunXMPP, a custom and proprietary modification of the XMPP protocol discussed in section 6.3. One of the reasons for the unprecedented success of WhatsApp has been the low barrier for entry : no need to have username and password registration, you can just start with a valid phone number. Over the years WhatsApp has added many features including audio-video calls and end-to-end encryption (see 8.1).

WhatsApp uses a highly optimized version of XMPP, which can be used in poor network environment, guaranteeing eventual (though ordering may differ) delivery of messages. This has led to its immense popularity in countries such as India with insufficient network infrastructure.

## 7.5 iMessage

Apple introduced iMessage with iPhone. iMessage seamlessly merges SMS and messages over Internet to create an illusion of unified inbox. This idea became very popular and was later used by Skype, Facebook Messenger and Google Hangouts.

# 8 Privacy and Security

## 8.1 End-to-End Encryption

# 9 Extensions

# 10 Future Work

# 11 Conclusion

newayz

# References


[1] "Cell phone usage statistics." https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/.

[2] E. A. Akkoyunlu, K. Ekanadham, and R. V. Huber, "Some constraints and tradeoffs in the design of network communications," *SIGOPS Oper. Syst. Rev.*, vol. 9, pp. 67–74, Nov. 1975.

[3] R. W. Burns, *Communications: An International History of the Formative Years*. IET, 2004.

[4] T. Standage, *The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century's On-line Pioneers*. Macmillan, 1998.

[5] G. L. Bodic, *Mobile Messaging Technologies and Services: SMS, EMS and MMS*. John Wiley & Sons, 2003.

[6] "The unidcode universal coded character set." http://www.columbia.edu/kermit/ucs2.html.

[7] "Sms point to point specification, 3gpp ts 23.040." https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=747.

[8] "The world in 2010.." http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf.

[9] "Unicode emoji list." http://unicode.org/emoji/charts/full-emoji-list.html.

[10] "Oxford 2015 word of the year.." http://blog.oxforddictionaries.com/2015/11/word-of-the-year-2015-emoji/.

[11] "imelody format specification.." http://merwin.bespin.org/t4a/specs/ems_imelody.txt.

[12] C. Erlandson and P. Ocklind, "Wap—the wireless application protocol," in *Mobile Networking with WAP*, pp. 165–173, Springer, 2000.

[13] D. Crocker, "Standard for the format of arpa internet text messages," 1982.

[14] P. W. Resnick, "Internet message format," 2008.

[15] H. Schulzrinne, "Real time streaming protocol (rtsp)," 1998.

[16] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "Real-time transport protocol," *RFC1899*, 2003.

[17] A. Bhushan, K. Pogran, R. Tomlinson, and J. White, "Standardizing network mail headers," tech. rep., 1973.

[18] D. Crocker, J. Vittal, K. T. Pogran, and D. A. Henderson, "Standard for the format of arpa network text messages," tech. rep., 1977.

[19] C. Kalt, "Internet relay chat: Architecture," tech. rep., 2000.

[20] C. Kalt, "Internet relay chat: Channel management," 2000.

[21] C. Kalt, "Internet relay chat: Client protocol," 2000.

[22] C. Kalt, "Internet relay chat: Server protocol," 2000.

[23] J. Oikarinen and D. Reed, "Internet relay chat protocol," 1993.

[24] P. Saint-Andre, "Extensible messaging and presence protocol (xmpp): Core," 2011.

[25] P. Saint-Andre, "Extensible messaging and presence protocol (xmpp): Instant messaging and presence," tech. rep., 2004.

[26] P. Saint-Andre, "Extensible messaging and presence protocol (xmpp): Address format," 2011.

[27] "Open system for communication in realtime - unofficial documentation." http://iserverd.khstu.ru/oscar/.

[28] "Skype moving to servers for scalability." https://www.listbox.com/member/archive/247/2013/06/sort/time_rev/page/1/entry/6:271/20130623090855:0B714E0A-DC06-11E2-9F35-8CD4CCA160A2/.

[29] "Messenger lite." FB newsroom.

[30] "Facebook messenger platform." https://developers.facebook.com/docs/messenger-platform.

[31] "Information technology – message queuing telemetry transport (mqtt) v3.1.1." http://www.iso.org/iso/catalogue_detail.htm?csnumber=69466.

[32] I. Fette, "The websocket protocol," 2011.

[33] "Whatsapp statistics." http://venturebeat.com/2016/02/01/whatsapp-passes-1-billion-monthly-active-users/.