Big Data Analysis of News Toolization Evolution Based on Computer Network Security Algorithms

Shengxuan Lin
College of Film, Television & Media,Guangxi Arts Institute , Nanning,Guangxi,53000, China
Email: YSGX345sx@163.com

Abstract:

In order to explore the inevitability of the evolution of the instrumentalization of data news, this paper combines the big data technology to study the inevitability of the evolution of the instrumentalization of news in the intelligent information age. Based on the MPTCP aggregate throughput model, this paper proposes an MPTCP throughput-aware traffic allocation and data scheduling algorithm (MPTCP-GA). Among them, the traffic allocation algorithm maximizes the MPTCP aggregate throughput through reasonable traffic allocation, and the data scheduling algorithm performs reasonable data distribution according to the results of traffic allocation. Moreover, this paper compares the proposed MPTCP-GA algorithm with CMT-QA and MPTCP round-robin scheduling algorithm through NS3 network simulator. The experimental research shows that the news big data system proposed in this paper can effectively improve the news data processing effect, which also verifies that big data tools have certain performance in news applications, and the evolution of data news tools is inevitable.

Keywords: big data; news application; data news; instrumentalization; evolution; network security

1 INTRODUCTION

Under the circumstance that online media holds a large amount of user data, the identity "weight" of traditional media content providers increases. In order to attract more time investment from the audience, the best way out for traditional media is to take advantage of the media convergence trend to give full play to their professional advantages, and strive to make breakthroughs in the depth and speed of dissemination of content. In addition to the acceleration of the big data era, data news is also the product of media integration, which can be regarded as the integration of media content production and big data technology [1]. Media integration has improved the status of professional media content production, and data journalism relies on highly credible big data technology to produce news content. In essence, it is for professional media to gain a solid place in media fusion. At the same time, big data can not only analyze the facts that have occurred, but also make predictions about the facts that have not occurred from the facts that have occurred, which is extremely important for professional media to grasp the opportunity of communication [2].

In the era of Web 2.0, social media has unique advantages in news communication. First of all, the development of Internet technology and the wide application of smart phones have cultivated new information acquisition habits of the audience, making online media the main channel for people to obtain information today, while social media that mainly rely on smart phone clients in online media get more traffic [3].

The new habit of obtaining information has given social media a dominant position. Secondly, social media has certain platform advantages in online news communication. The public are often the first witnesses of news events, and have the most authentic, direct and first-hand resources. The low threshold of social media and the convenience of online information release make the status of the first publisher of news events snatched by Internet users, while the release platform has changed from traditional media to social media [4]. Thirdly, compared with newspapers, magazines, television and other media, the content released by social media can be corrected immediately according to readers' feedback, and can also achieve two-way interaction through messages, comments, sharing and so on. Finally, through the cultivation of audience habits for a long time, social media has accumulated a large number of user resources, making the number of audiences larger and the status more stable. The communication content and mode have also been optimized through continuous adaptation, gradually finding their own position, and the development is more stable [5]. In addition, netizens have received a large amount of information with rich content through the Internet, which has broadened their horizons and increased their knowledge. They have awakened their own right to speak and their sense of independence to varying degrees. The most obvious manifestation is that grassroots Internet celebrities continue to emerge. And with the increase of netizens' time in contact with the Internet, the media literacy of the audience has been greatly improved compared with the previous years, especially in recent years, the reverse news has emerged frequently, The audience's wait-and-see attitude has become obvious, and they are not eager to draw conclusions about news events. At the same time, the proportion of pertinent and rational comments in various comments and messages is also increasing. Even through background screening, we can see the rational regression trend of current Internet users' online behavior [6].

The continuous innovation of Internet, big data, cloud computing and other technologies has pushed the society into the age of big data. Information dissemination in all walks of life and in all fields is increasingly showing the characteristics of data, visualization and visualization. With the continuous progress of data visualization technology, the breadth and depth of data applications have been improved to varying degrees, and data collection has become more efficient with the help of cloud and data center [7]. The application of big data in the field of news communication has gradually generated data news. In a sense, big data expands the intrinsic meaning of news, changes the way of news communication, optimizes the communication effect of news and improves the convenience of news processing. At the same time, with the deepening of academic and industry research on data news, the understanding and utilization efficiency of data have also been improved [8]. Data is no longer just evidence to help the audience understand the news, or media workers find news clues through the statistics of surface data. Many scholars realize that the meaning of data itself is a kind of information, and its information content is far beyond what they see on the surface. More and more practitioners and researchers also realize that using the correlation of data can predict the content that is easy to be ignored but has important value [9].

The network is already an important medium of mass communication. First of all, there are not only various websites and media organizations engaged in news communication and mass information services, but also a large number of non professional information publishers derived from the low threshold of online information release. The convenience and real-time nature of the network make information disseminators not only limited to professional journalists, but also allow any eyewitness on the scene of news events to spread information to the outside world through the network. This has had a significant impact on professional media organizations [10]. Before the emergence of the "national journalists", the news communication process was often first reported by professional media, and then spread among the public. However, nowadays, Internet users and professional media are on the same footing, playing the role of communicators, while the media and the public have become the audience in the communication process, and the boundaries between communicators and receivers are becoming increasingly blurred. Secondly, information productivity continues to create new social achievements. Network information platforms provide more and more services, which constantly distract, divert and consume people's attention. Knowledge, time and work are fragmented. At the same time, the increasingly refined social division of labor also accelerates the process of fragmentation. Fragmentation has become a synonym of this era [11]. Under such characteristics of the times, it is extremely important for the news industry to attract the audience's distracted attention in the complex information age. Whether it is the voice of "content is king" or "service is king", it is inseparable from the design of news content. Nowadays, retaining audiences with pictures has become a thing of the past. Expression packs, micro videos, live broadcasts, etc. are gradually enriching the expression of news content, and the continuous improvement of data visualization makes it possible for the audience to simply interpret obscure depth data [12]. With the development of We Media and the lowering of the threshold for Internet voice, Internet users have awakened their self-consciousness, and their subjective initiative has been highlighted. Their attention to their own voice has reached an unprecedented height. The most prominent performance is the emergence of a large number of grassroots Internet celebrities. People realize that the Internet gives everyone a stage to show, and ordinary people also have the opportunity to become famous. At the same time, people openly express their opinions on the network, actively share the anecdotes happening around them or forward interesting network information, gradually realize their position in network communication is constantly improving, and begin to actively participate in the first end of network information communication [13]. In the network information dissemination, the netizens at the scene of news events are often the news reporters. They have more first-hand information than professional media, and they are more willing to speak out about the truth. More and more netizens are also willing to spread the event scene through their own platform accounts. Internet users are more and more active in spreading information on the network, which reflects that Internet users are not only satisfied with making a simple voice on the network, but are more willing to participate in the network information dissemination with a positive attitude relying on their own resources, and become a more professional network communication subject [14].

Today, the rapid development of the Internet has brought convenience to news dissemination, but at the same time, a series of problems have emerged. The main force of network information communication has changed from media to Internet users and media. The growth of Internet users directly leads to the media taking the audience first in the agenda setting of network information content. The most obvious problem is the homogeneity of communication content [15]. The causes of this problem are mainly as follows: First, the network information is in the normal state of explosive growth, news content is updated in real time, and hot spots change frequently. At the same time, the network information environment is full of many useless, homogeneous junk information, which makes the media tired of coping with the time race, and lacks originality and deep thinking in the push of news content [16]. Second, audience oriented news topics are deeply rooted in media editing awareness. In order to compete for more traffic and attract more people's attention in the network environment, the information content pays too much attention to the title, topics enlarge the audience's needs, superstitious click through rate, and blindly cater to the audience. However, most Internet users form a fragmented reading habit in the context of new media, and the sensory enjoyment of news content is gradually replacing deep thinking, resulting in the news content spread by the media on the way of chasing the audience becoming more and more superficial, Gradually tend to homogenize [17]. Third, the rapid development of network technology continues to produce new forms of communication. Some media blindly follow the trend of new forms of communication without considering creating products that conform to their own communication laws. Often, new bottles of old wine are filled with new ones, and the soup is not changed, leading to more serious homogenization of the content of new media [18]

This paper combines big data technology to study the inevitability of the evolution of news in the intelligent information age, and uses the system model to verify the model effect and explore the development path of intelligent news.

2 MPTCP NEWS TRAFFIC DISTRIBUTION AND NEWS DATA SCHEDULING ALGORITHM

2.1 System model

The system block diagram is shown in Figure 1. The MPTCP transmission framework includes a sender, a receiver and P sub-streams.

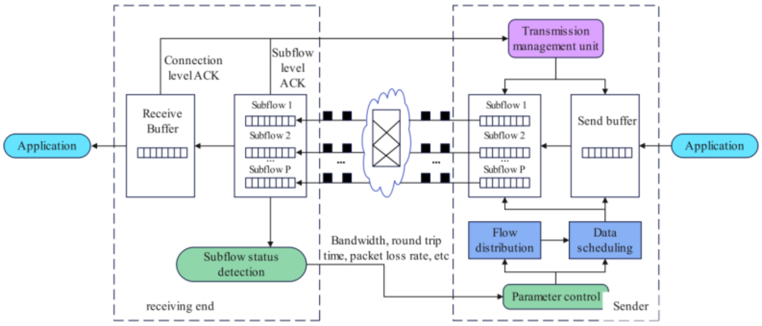


Figure 1 MPTCP system block diagram

Each subflow in MPTCP is modeled as a separate communication path $j, j \in \mathrm{P}, \mathrm{P}$ is the set of subflows. We assume that the sub-streams are independent of each other, and use the bandwidth μ_j, round-trip time RTT_j and packet loss rate π_j^B to characterize the status of each sub-stream.

The transition probability from state B to state G is ξ_j^B, and the transition probability from state G to state B is ξ_j^G. Two parameters are used to characterize the packet loss model of this continuous-time Markov process, the average burst packet loss duration $\frac{1}{\xi_j^B}$ and the path packet loss rate $\pi_j^B = \frac{\xi_j^B}{\xi_j^B + \xi_j^G}$.

The time from when the packet is scheduled to sub-stream j to the time it is retransmitted to the receiver after a timeout is:

$$d_j^{TO} = min\left(RTO_j + \frac{RTT_j}{2}, RTO_j + E\left[RTT_k\right] \right) \quad (1)$$

When choosing another sub-stream k, we adopt a simple strategy to preferentially select the sub-stream with the smallest packet loss rate to retransmit the lost news data packets. The time-out news data packet can be successfully sent to the receiver through the sub-stream with low packet loss rate. Therefore, $E\left[RTT_k\right]$ can be expressed as:

$$E\left[RTT_k\right] = \sum_{k=1,k\neq j}^{k=P} \frac{1/\pi_k^B}{\sum_{l=1,l\neq j}^{l=P} 1/\pi_k^B} * RTT_k \quad (2)$$

P sub-streams in MPTCP share the same receive buffer. When the receive buffer is infinitely large, the aggregate throughput of MPTCP is:

$$E[TP] = \sum_{j=1}^{j=P} E\left[TP_j\right] \quad (3)$$

The algorithm models finite receive buffer MPTCP aggregate throughput. The size of the receive buffer is L_{Buf} , the number of MPTCP sub-streams is P, and the throughput of each sub-stream is $E\left[TP_j\right]$. we consider the receive buffer to be blocked. $P(j,i)$ denotes the probability that sub-stream j is not blocked by news data packets transmitted on sub-stream i, and $P\left(j,^*\right)$ denotes the probability that sub-stream j is not blocked by any other sub-stream. The aggregate throughput of MPTCP can be expressed as:

$$E[TP] = \sum_{j=1}^{j=P} P\left(j,^*\right) * E\left[TP_j\right] \quad (4)$$

First, the modeling analysis is carried out through the MPTCP scenario with only two sub-streams, and then it is extended to the scenario of multiple sub-streams.

(1) Two sub-streams

Without loss of generality, we assume that the sub-stream with larger RTT among the two sub-streams is sub-stream 1, and set $l = \frac{RTT_1}{RTT_2} \geq 1$. When the packet s_1 is sent through sub-stream 1 and has not yet reached the receiver, the f packets transmitted by the algorithm through sub-stream 2 arrive at the receive buffer. If $n_2 \geq L_{Buf}$, the receive buffer is full, sub-stream 2 can no longer send packets, and sub-stream 2 is blocked by sub-stream 1. r represents the number of rounds required to fill the receive buffer sub-stream 2 (one round time is equal to one RTT_2 time), and r can be expressed as:

$$r = \frac{L_{Buf}}{RTT_2 * E\left[TP_2\right]} \quad (5)$$

The probability of sub-stream i entering the timeout retransmission state after packet loss occurs can be expressed as:

$$Q_i = 1-\left(1-\pi_i^B\right)^3 \quad (6)$$

When the ratio of RTT l is different, the transmission of sub-stream 2 may be blocked in these three cases. The following three cases are analyzed separately.

Case1: Substream 1 transmits packet s_1 to the receive buffer over time $\frac{RTT_1}{2}$ with probability $1-\pi_1^B$. If $\frac{RTT_1}{2} \geq r^* RTT_2$, even if no packet loss occurs during the transmission of sub-stream 1, the packet successfully reaches the receive buffer, which will block the transmission of sub-stream 2. We define:

$$(a)^+ = \begin{cases} a & a \leq 1 \\ 1 & a > 1 \end{cases} \quad (7)$$

The probability that sub-stream 1 is in this situation and does not block the transmission of sub-stream 2 is:

$$P_{casel}(2,1) = \left(1-\pi_1^B\right) * \left(\frac{r^* RTT_2}{RTT_1 / 2}\right)^+ = \left(1-\pi_1^B\right)\left(\frac{2r}{l}\right)^+ \quad (8)$$

Case2: The packet is lost with probability $\pi_1^B * \left(1-Q_1\right)$ and a timeout retransmission is triggered.

If l<2r, the news data packets transmitted by sub-stream 1 are not lost, and the transmission of sub-stream 2 will not be blocked.

The time from when a news data packet is sent to reach the receive buffer through fast retransmission is about $\frac{3RTT_1}{2}$. The probability that sub-stream 1 is in this situation and does not block the transmission of sub-stream 2 is:

$$P_{case2}(2,1) = \pi_1^B * \left(1-Q_1\right) * \left(\frac{r^* RTT_2}{3RTT_1 / 2}\right)^+ = \pi_1^B \left(1-Q_1\right)\left(\frac{2r}{3l}\right)^+ \quad (9)$$

Case3: The packet is lost with probability $\pi_1^B * Q_1$ and a timeout retransmission is triggered.

If $l < \frac{2r}{3}$, when sub-stream 1 is in both Case1 and Case2 conditions, it will not block the transmission of sub-stream 2.

According to the analysis of the timeout retransmission model, the time from the beginning of the news packet is sent to the time it reaches the receive buffer through the timeout retransmission is about d_1^{TO}. The probability that sub-stream 1 is in this situation and does not block the transmission of sub-stream 2 is:

$$P_{case3}(2,1) = \pi_1^B * Q_1 * \left(\frac{r^* RTT_2}{d_1^{TO}}\right)^+ \quad (10)$$

Combining the above three situations, the probability that sub-stream 2 is not blocked by sub-stream 1 is:

$$P(2,1) = P_{casel}(2,1) + P_{case2}(2,1) + P_{case3}(2,1) \quad (11)$$

When there are two sub-streams, the probability $P\left(2,^{*}\right)$ that sub-stream 2 is not blocked by any sub-stream is equal to the probability P(2,1) that sub-stream 2 is not blocked by sub-stream 1, that is, P(2,*)=P(2,1).

Similarly, the probability $P\left(1,^{*}\right)$ that sub-stream 1 is not blocked by other sub-streams can also be obtained. By substituting P(1,*) and P(2,*) into Equation (4), the aggregate throughput of MPTCP with two sub-streams is obtained.

(2) Multiple sub-streams

When there are P sub-streams in an MPTCP connection, the RTTs of these sub-streams satisfy:

$$l_1 : l_2 : \ldots : l_P = RTT_1 : RTT_2 : \ldots : RTT_P \quad (12)$$

First, the algorithm derives the probability P(j,i) that sub-stream j is not blocked by sub-stream i. Before the packets of sub-stream i arrive at the receiving buffer, the news data packets sent by P-1 sub-streams except sub-stream i may all reach the receiving buffer. Therefore, when analyzing the number of rounds r required for the receive buffer to be filled, other P-1 sub-streams need to be considered together. r is re-expressed as:

$$r = \frac{L_{Buf}}{\left(\sum_{k=1,k\neq i}^{k=P} \frac{l_j}{l_k} * E\left[TP_k\right]^{*} RTT_k\right)} \quad (13)$$

Extending formula (11), the probability P(j,i) that sub-stream j is not blocked by sub-stream i is:

$$\begin{aligned} P(j,i) = &\left(1-\pi_i^B\right) * \left(\frac{2rl_j}{l_i}\right)^{+} + \pi_i^B * \left(1-Q_i\right) * \left(\frac{2rl_j}{3l_i}\right)^{+} \\ &+ \pi_i^B * Q_i * \left(\frac{r^{*} RTT_j}{d_i^{TO}}\right)^{+} \end{aligned} \quad (14)$$

The probability $P\left(j,^{*}\right)$ that sub-stream j is not blocked by any sub-stream is:

$$P\left(j,^{*}\right) = 1 - \sum_{i=1,i\neq j}^{i=P} \left[(1-P(j,i))^{*} \prod_{k\neq i,k\neq j} P(j,k) \right] \quad (15)$$

When the news flow R_j allocated by the sub-stream satisfies $R_j \le \mu_j$, where μ_j is the available bandwidth of the sub-stream j, the throughput of the MPTCP sub-stream can be expressed as:

$$E\left[TP_j\right] = E\left[R_j\right] * \left(1-\pi_j^B\right) \quad (16)$$

2.2 MPTCP-GA news traffic distribution algorithm

Combined with the sub-stream parameter model, the timeout retransmission model, the limited receive buffer model and the sub-stream throughput model, it is assumed that there are P sub-streams (represented by the set P) in MPTCP. If the news flow allocated by the j-th subflow is denoted as R_j, the aggregate throughput of MPTCP can be expressed as:

$$E[TP] = \sum_{j=1}^{j=P} P(j,*) * E\left[R_j\right] * \left(1-\pi_j^B\right) \quad (17)$$

In a news flow distribution period, the news flow allocated to each sub-stream is a fixed value. Therefore,

$E\left[R_j\right]=R_j$.

The aggregate throughput of MPTCP can be expressed as:

$$E[TP]=\sum_{j=1}^{j=P}P\left(j,^*\right)*R_j*\left(1-\pi_j^B\right) \quad (18)$$

The optimization problem can be modeled as:

$$max_{\{R\}}\{E[TP]\} \quad (19)$$

$$s.t.\left(C_1\right)0\le R_j\le\mu_j,\forall j\in \mathrm{P}$$

$$\left(C_2\right)E[TP]=\sum_{j=1}^{j=P}P\left(j,^*\right)*R_j*\left(1-\pi_j^B\right)$$

$$\left(C_3\right)P\left(j,^*\right)=1-\sum_{i=1,1\neq j}^{i=P}\left[(1-P(j,i))^*\prod_{k\neq i,k\neq j}P(j,k)\right]$$

$$\left(C_4\right)P(j,i)=\left(1-\pi_i^B\right)^*\left(\frac{2rl_j}{l_i}\right)^+ +\pi_i^B*\left(1-Q_i\right)^*\left(\frac{2rl_j}{3l_i}\right)^+$$

$$+\pi_i^B*Q_i*\left(\frac{r*RTT_j}{d_i^{TO}}\right)^+$$

$$\left(C_5\right)r=\frac{L_{Buf}}{\left(\sum_{k=1,k\neq i}^{k=P}\frac{l_j}{}l_k*R_k*\left(1-\pi_k^B\right)^*RTT_k\right)}$$

$$\left(C_6\right)l_1:l_2:\ldots:l_P=RTT_1:RTT_2:\ldots:RTT_P$$

$$\left(C_7\right)Q_i=1-\left(1-\pi_i^B\right)^3$$

We assume that the objective function f(x) is differentiable near x_k, and $g_k @\nabla f\left(x_k\right)\neq 0$. f(x) is Taylor expanded to $f(x)=f\left(x_k\right)+g_k^T\left(x-x_k\right)+o\left(\left\|x-x_k\right\|\right)$ at x_k. If $x-x_k=\alpha d_k$, the above formula can be expressed as $f\left(x_k+\alpha d_k\right)=f\left(x_k\right)+\alpha g_k^T d_k+o\left(\left\|\alpha d_k\right\|\right)$.

Since the objective problem (19) is a maximization problem, the inversion of the objective function of this problem is transformed into a minimization problem, and the objective function is redefined as:

$$\psi=-\sum_{j=1}^{P}P\left(j,,^*\right)*R_j*\left(1-\pi_j^B\right) \quad (20)$$

The maximum value of the objective problem (19) is converted into finding the minimum value of ψ. First, the algorithm finds the gradient vector of ψ.

$$\mathrm{g}=\left[\frac{\partial\psi}{\partial R_1},\frac{\partial\psi}{\partial R_2},\ldots,\frac{\partial\psi}{\partial R_P}\right] \quad (21)$$

Among them,

$$\frac{\partial\psi f}{\partial R_j}=-\sum_{n=1,n\neq j}^{P}R_n\left(1-\pi_n^D\right)\frac{\partial P\left(n,^*\right)}{\partial R_j}-\left(1-\pi_j^D\right)P\left(j,^*\right)-\left(1-\pi_j^D\right)R_j\frac{\partial P\left(j,^*\right)}{\partial R_j} \quad (22)$$

To do this, we need to find $\frac{\partial P(n,*)}{\partial R_j}$. By transforming formula (15), we get:

$$P\left(n,^*\right)=1-\sum_{i=1,i\neq n}^{P}\prod_{k\neq i,k\neq n}P(n,k)+(P-1)\prod_{k\neq n}P(n,k) \quad (23)$$

The algorithm derives from $\prod_{k\neq i,k\neq n}P(n,k)$ and $\prod_{k\neq n}P(n,k)$ respectively to get:

$$\frac{\partial\left(\prod_{k\neq i,n}P(n,k)\right)}{\partial R_j}=\sum_{s=1,s\neq i,n}^{P}\frac{\partial P(n,s)}{\partial R_j}\prod_{k\neq i,n,s}P(n,k) \quad (24)$$

$$\frac{\partial\left(\prod_{k\neq n}P(n,k)\right)}{\partial R_j}=\sum_{s=1,s\neq n}^{P}\frac{\partial P(n,s)}{\partial R_j}\prod_{k\neq n,s}P(n,k) \quad (25)$$

The algorithm substitutes formulas (24) and (25) into formulas (23) to get:

$$\frac{\partial P\left(n,^*\right)}{\partial R_j}=-\sum_{i=1,i\neq n}^{P}\left(\sum_{s=1,s\neq i,n}^{P}\frac{\partial P(n,s)}{\partial R_j}\prod_{k\neq i,n,s}P(n,k)\right)+(P-1)\left(\sum_{s=1,s\neq n}^{P}\frac{\partial P(n,s)}{\partial R_j}\prod_{k\neq n,s}P(n,k)\right) \quad (26)$$

Similarly, we get:

$$\frac{\partial P\left(j,^*\right)}{\partial R_j}=-\sum_{i=1,i\neq j}^{P}\left(\sum_{s=1,,\neq i,j}^{P}\frac{\partial P(j,s)}{\partial R_j}\prod_{k\neq i,j,s}P(j,k)\right)+(P-1)\left(\sum_{s=1,s\neq j}^{P}\frac{\partial P(j,s)}{\partial R_j}\prod_{k\neq j,s}P(j,k)\right) \quad (27)$$

Since P(n,s) contains the form of $(a)^+$, it is not convenient to directly describe the derivative of P(n,s) with respect to R_j. However, in each calculation process, when the news flow distribution vector is at a certain point, the value of r can be obtained by formula (13). Then, the $(a)^+$ form can be removed by substituting r into P(n,s). At this point, the algorithm takes the derivative of P(n, s) to obtain $\frac{\partial P(n,s)}{\partial R_j}$, and substituting formulas (26) and (27) into formula (22) can obtain $\frac{\partial\psi}{\partial R_j}$.

2.3 MPTCP-GA news data scheduling algorithm

In order to cooperate with the MPTCP-GA news traffic distribution algorithm and alleviate the problem of increasing packet transmission delay caused by out-of-order packets in the receiving buffer, this section proposes the MPTCP-GA news data scheduling algorithm. The purpose is to make news data packets arrive in the receiving buffer in sequence, and further improve the transmission performance of MPTCP.

The algorithm determines the time interval vector $\tau = [\tau_1, \tau_2, ..., \tau_P]$, sets the scheduling timer for each sub-stream according to the time interval vector τ, and schedules the news data packets in the sending buffer to the sending buffer of the sub-stream. That is, one news data packet is scheduled to sub-stream j every τ_j time. The time interval τ_j for news data scheduling for sub-stream j can be expressed as:

$$\tau_j = \frac{MSS_j}{R_j} \quad (28)$$

The algorithm marks the unscheduled news data packet with the smallest sequence number in the sending buffer of the sender as s_{min}. When scheduling news data packets for the jth sub-stream, if s_{min} is scheduled to this sub-stream, the number of out-of-order packets generated by the receive buffer can be expressed as:

$$N_j = \frac{\sum_{\substack{i \in P \\ RTT_i < RTT_j}} R_i * \left(1 - \pi_i^B\right) * \left(RTT_j - RTT_i\right) / 2}{MSS_j} \quad (29)$$

2.4 Simulation results and analysis

This section uses the heterogeneous wireless network environment shown in Figure 2 as the simulation scenario. As the sending end, the server sends video files to the multi-mode terminal through three sub-streams, wherein the detailed parameter settings of the three sub-streams are shown in Table 1.

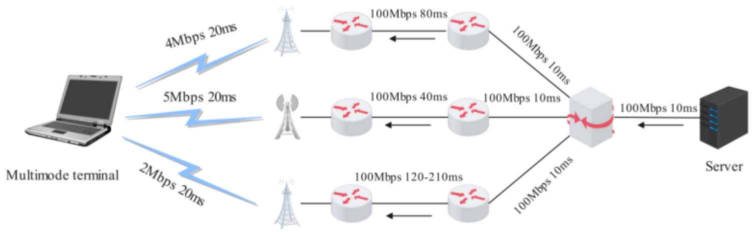


Figure 2 Simulation network topology

As can be seen from Figure 3, the throughput of MPTCP-GA reaches the maximum value when the receive buffer size is 128KB, and the throughput of CMT-QA and MPTCP round-robin scheduling algorithm reaches the maximum value when the receive buffer size is 256KB.

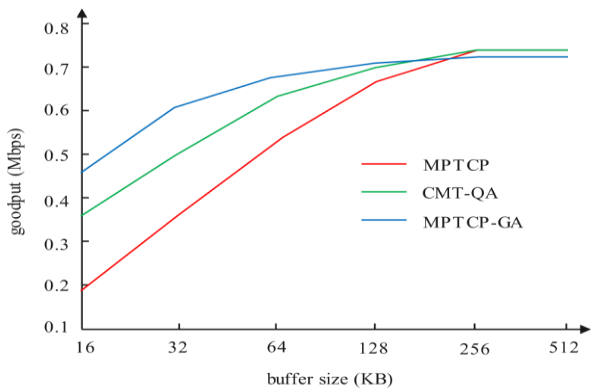


Figure 3 The aggregate throughput of MPTCP-GA, MPTCP and CMT-QA varies with the size of the receive buffer

According to Figure 3, when the receiving buffer is large enough, the receiving buffer can accommodate enough out-of-order news data packets, and the importance of news flow distribution and news data scheduling cannot be well highlighted. The receive buffer size is set to 32KB, and the throughput of the three schemes and their sub-stream throughput changes with time are observed. The results are shown in Figure 4. Figures 4(a)~(d) depict the theoretical results obtained by solving the optimization problem (19) proposed in this chapter, the throughput variation diagrams of the MPTCP-GA scheme, the MPTCP round-robin scheduling scheme and the CMT-QA scheme, respectively. The abscissa is time and the ordinate is throughput.

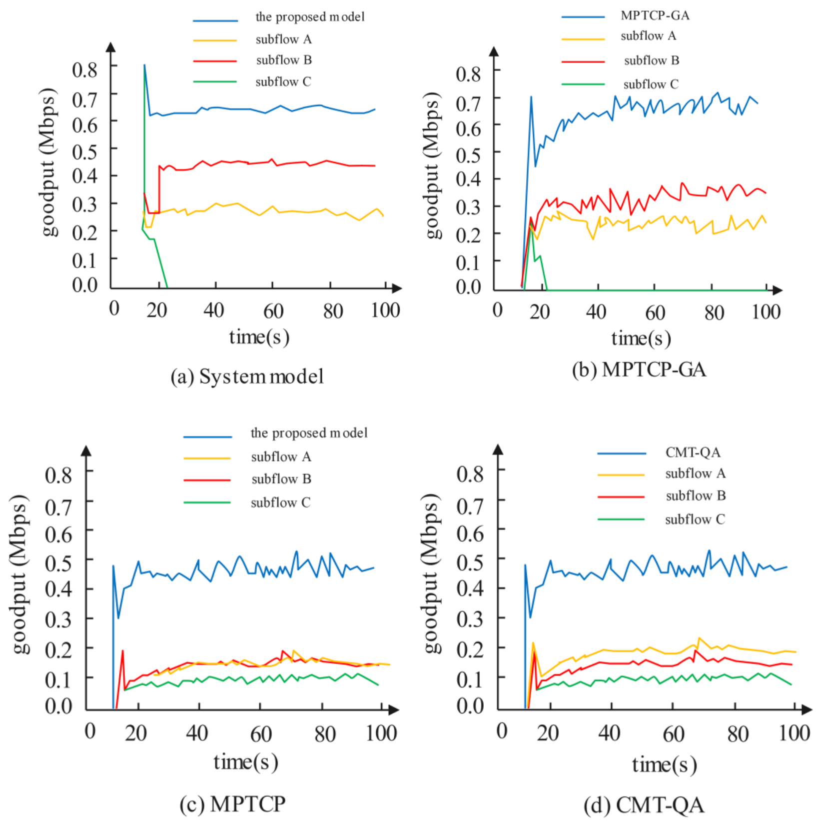
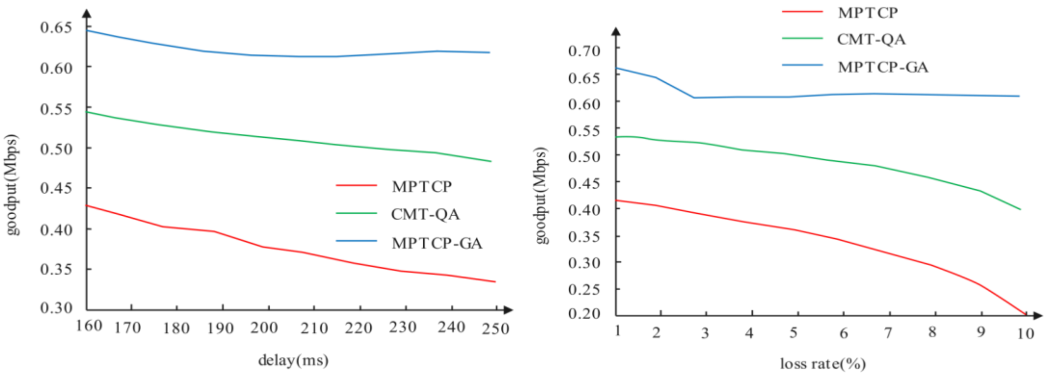


Figure 4 The throughput of system model, MPTCP-GA, MPTCP, CMT-QA with the change of time

Figure 4(a) and (b) show that when MPTCP-GA realizes that sub-stream C will hinder the transmission of sub-stream A and sub-stream B, it no longer allocates news traffic to sub-stream C, which effectively avoids the problem of blocking the receive buffer caused by sub-stream differences. Figure 4(c) shows that the throughputs of each sub-stream in MPTCP are relatively similar. Under the condition of limited receive buffer, sub-stream C with poor performance affects the transmission of sub-stream A and sub-stream B, which reduces the aggregate throughput of MPTCP. Figure 4(d) shows that the throughput of sub-stream B is the highest in CMT-QA, followed by sub-stream A, and sub-stream C is the lowest.

To further observe the effect of sub-stream differences on MPTCP-GA, we transmit a 20MB video file in the network topology shown in Figure 2. Moreover, we change the delay and packet loss rate of sub-stream C respectively, and observe the performance of MPTCP-GA and the comparison scheme in terms of average throughput, average packet transmission delay, etc., and obtain the results shown in Figure 5 and Figure 6, respectively.

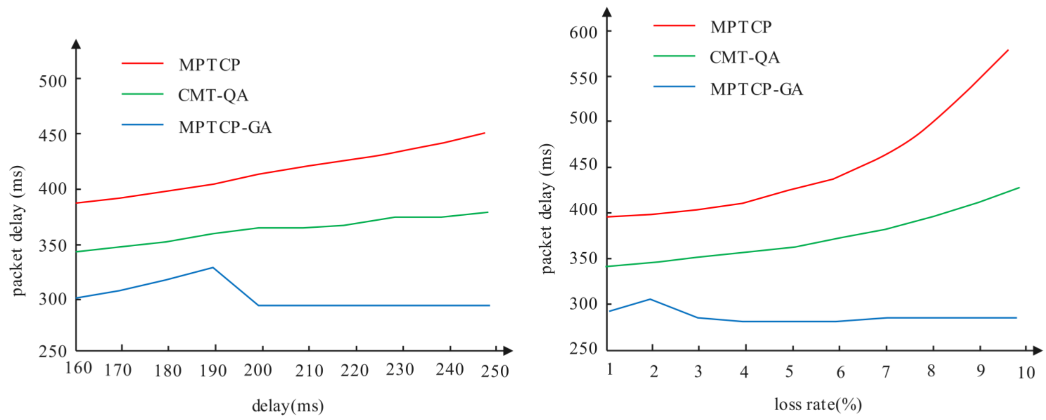


(a) Time delay (b) Packet loss rate

Figure 5 The aggregate throughput of MPTCP-GA, MPTCP, and CMT-QA varies with the delay and packet loss rate of subflow C

Figure 5 shows the relationship between the throughput of the three schemes and the delay and packet loss rate of sub-stream C.

Figure 6 shows the relationship between the average transmission delay of news data packets of the three schemes and the delay and packet loss rate of sub-stream C.



(a) Time delay (b) Packet loss rate

Figure 6 The average packet transmission delay of MPTCP-GA, MPTCP, and CMT-QA varies with the delay and packet loss rate of sub-stream C

It can be observed from Figure 7 that the number of out-of-order packets in the receive buffer of MPTCP-GA is generally small, and the receive buffer is not blocked. In CMT-QA, the receive buffer is blocked once, while the receive buffer of the MPTCP round-robin scheduling scheme has a large number of out-of-order packets, and the receive buffer is blocked many times.

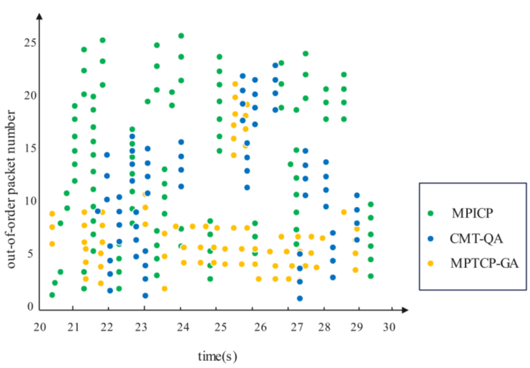


Figure 7 The number of out-of-order packets in the receive buffer of MPTCP-GA, MPTCP, and CMT-QA varies with the change of time

3 THE INEVITABILITY AND NECESSITY OF THE INSTRUMENTALIZATION EVOLUTION OF DATA JOURNALISM BASED ON BIG DATA TOOLS

The news big data system adopts distributed big data technology frameworks such as Hadoop and Spark in the overall architecture. Among them, the BI system is built on the Hadoop cluster, providing visual analysis tools and data mining models, etc. The data storage and data processing use Hadoop's distributed file system HDFS and distributed parallel computing framework MapReduce, as shown in Figure 8.

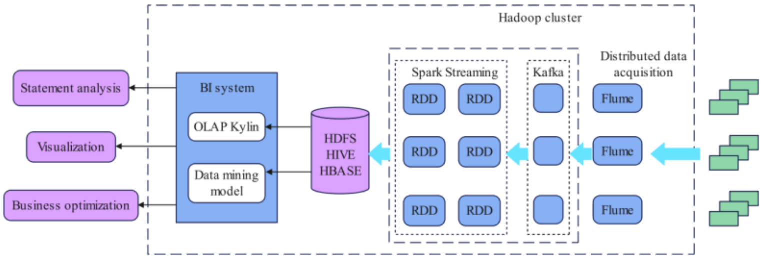


Figure 8 Architecture of news big data system

The pre-trained news event detection model is used to construct a multi-label data set for news event prediction, and the structure frame of the symptom event prediction model is shown in Figure 9.

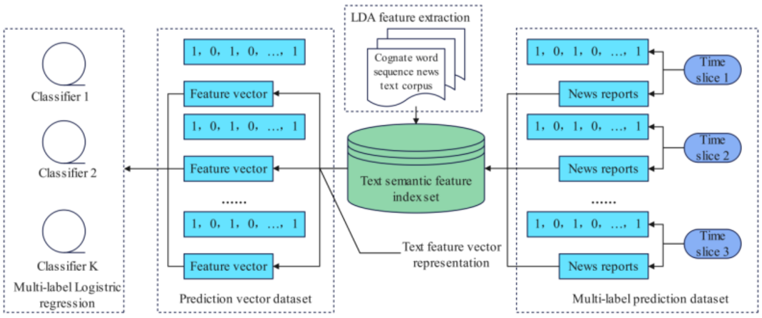


Figure 9 Structure box of news event prediction model

The effect of the news big data system proposed in this paper is verified, and the effect of the model in news data processing is counted, and the results are shown in Table 1.

Table 1 Verification of the effect of news big data system

Number	News processing	Number	News processing
1	82.4721	14	84.0439
2	88.8467	15	81.1293
3	86.3457	16	84.2441
4	81.6049	17	88.3271
5	83.4693	18	85.0029
6	81.3372	19	84.3201
7	87.9529	20	81.9945
8	86.6564	21	86.1636
9	82.6642	22	85.2994
10	84.1526	23	85.9399
11	87.2482	24	82.2833
12	85.0500	25	88.7996
13	83.6001	26	87.1189

From the above research, it can be seen that the news big data system proposed in this paper can effectively improve the effect of news data processing, which also verifies that big data tools have a certain performance in news applications, and the evolution of data news tools is inevitable.

4 CONCLUSION

With the continuous deepening of the research on network media in the academic circles, media fusion has been endowed with a new era significance and a new media output mode has been derived. At present, the practical content of media integration has covered content integration, platform integration, channel integration and model integration. The greatest impact of media convergence on communication is that traditional information dissemination channels are being integrated with the Internet, mobile Internet and social media applications. Moreover, this kind of "unequal" integration makes the communication channels mainly large-scale Internet platforms or mobile phone platforms with huge users and data resources. This paper combines the big data technology to study the inevitability of the evolution of news tools in the intelligent information age. The research shows that the news big data system proposed in this paper can effectively improve the effect of news data processing.

REFERENCES

[1] Kaur, S., Kumar, P., & Kumaraguru, P. (2020). Automating fake news detection system using multi-level voting model. Soft Computing, 24(12), 9049-9069.

[2] Marwick, A. E. (2018). Why do people share fake news? A sociotechnical model of media effects. Georgetown law technology review, 2(2), 474-512.

[3] Xu, D., Pan, J., Wang, B., Liu, M., & Kang, Q. (2019). Information dissemination model of microblogging with Internet marketers. Journal of Information Processing Systems, 15(4), 853-864.

[4] Kelly, M. P., Martin, N., Dillenburger, K., Kelly, A. N., & Miller, M. M. (2019). Spreading the news: History, successes, challenges and the ethics of effective dissemination. Behavior Analysis in Practice, 12(2), 440-451.

[5] Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political psychology in the digital (mis) information age: A model of news belief and sharing. Social Issues and Policy Review, 15(1), 84-113.

[6] Ghosh, S., & Shah, C. (2018). Towards automatic fake news classification. Proceedings of the Association for Information Science and Technology, 55(1), 805-807.

[7] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big data, 8(3), 171-188.

[8] Jadhav, S. S., & Thepade, S. D. (2019). Fake news identification and classification using DSSM and improved recurrent neural network classifier. Applied Artificial Intelligence, 33(12), 1058-1068.

[9] Huang, H., Wang, T., Hu, M., Dong, M., & Lai, L. (2021). Node attitude aware information dissemination model based on evolutionary game in social networks. Mobile Networks and Applications, 26(1), 114-129.

[10] Amini, A., & Ebrahimi Gamsari, F. (2019). Model of the Effect of Error Management Culture Dissemination on Quality of Work-life and Psychological Empowerment of Personnel. Transformation Management Journal, 11(2), 1-28.

[11] Behrooz, B., Delbari, A., & FayyazMohammadi, M. (2021). Modeling Oil Pollution Dissemination in Different Areas of Arvandroud River, Iran. Environmental Researches, 12(23), 185-198.

[12] Shrivastava, G., Kumar, P., Ojha, R. P., Srivastava, P. K., Mohan, S., & Srivastava, G. (2020). Defensive modeling of fake news through online social networks. IEEE Transactions on Computational Social Systems, 7(5), 1159-1167.

[13] Sugita, Y., Yamashita, K., Takimoto, R., Tanaka, T., Arimoto, A., Fukuoka, E., ... & Kakeji, Y. (2019). Intraperitoneal myeloid-derived suppressor cells in peritoneal dissemination mouse model. Cancer Research, 79(13_Supplement), 1527-1527.

[14] Shrivastava, G., Kumar, P., Ojha, R. P., Srivastava, P. K., Mohan, S., & Srivastava, G. (2020). Defensive modeling of fake news through online social networks. IEEE Transactions on Computational Social Systems, 7(5), 1159-1167.

[15] Borges-Tiago, T., Tiago, F., Silva, O., Guaita Martinez, J. M., & Botella-Carrubi, D. (2020). Online users' attitudes toward fake news: Implications for brand management. Psychology & Marketing, 37(9), 1171-1184.

[16] Xiao, X., Borah, P., & Su, Y. (2021). The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs. Public Understanding of Science, 30(8), 977-992.

[17] Torres, R., Gerhart, N., & Negahban, A. (2018). Epistemology in the era of fake news: An exploration of information verification behaviors among social networking site users. ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 49(3), 78-97.

[18] Nakayama, J., Lu, J. W., Makinoshima, H., & Gong, Z. (2020). A novel zebrafish model of metastasis identifies the HSD11β1 inhibitor adrenosterone as a suppressor of epithelial–mesenchymal transition and metastatic dissemination. Molecular Cancer Research, 18(3), 477-487.